

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 54.0796 Seconds
(without alignments)
239.005 Million cell updates/sec

Title: US-09-854-133-586
Perfect score: 532
Sequence: 1 EWEVSRDHASLGSETLSQT.....LTGGCLPWATSHLGRKCS 97

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	18.4	16	22	AAE13851
2	82	15.4	121	22	AAO13020
3	71.5	13.4	46	22	AAO04640
4	71.5	13.4	141	22	AAO09875
5	71.5	13.4	163	22	AAO75462
6	69.5	13.1	66	22	ABB03142
7	69	13.0	63	22	ABBI7957
8	68	12.8	423	22	AAO51507
9	68	12.8	423	22	AAE03740
10	68	12.8	423	23	AAE20598

11	68	12.8	458	20	AAV28851	Active form of hum
12	68	12.8	458	21	AAV90676	Human mutant G pro
13	68	12.8	458	22	AAAB45803	Human serotonin re
14	67.5	12.7	105	22	AAAB93643	Human ORFX protein
15	67	12.6	81	23	AAPO00516	Human secreted pro
16	67	12.6	190	22	AAAB4601	Human secretin va
17	67	12.6	458	17	AAAR94577	Human 5-HT2C serot
18	67	12.6	458	19	AAAR77108	Human G protein-co
19	67	12.6	458	21	AAV90641	Human reproductive
20	66.5	12.5	64	22	AAAG95632	Human polypeptide
21	66.5	12.5	116	22	AAO12091	Human polypeptide
22	66.5	12.5	138	22	AAO00532	Plasmodium falcipa
23	66.5	12.5	2197	21	AAAB18352	Novel human diagn
24	66	12.4	83	22	ABG26719	Novel human diagn
25	66	12.4	100	22	ABG11242	Common bean COR-4
26	66	12.4	369	23	AAW51135	Active form of hum
27	66	12.4	478	20	AAV28853	Human serotonin re
28	66	12.4	478	22	AAAB45805	Invertebrate octop
29	66	12.4	637	20	AAV13445	Drosophila melanog
30	66	12.4	637	22	ABB60216	D. melanogaster oc
31	66	12.4	642	20	AAV33680	Drosophila melanog
32	66	12.4	642	20	AAV25464	Drosophila melanog
33	65.5	12.3	381	22	ABG65092	Propionibacterium
34	65	12.2	232	21	AAO7926	Arabidopsis thalia
35	65	12.2	418	10	AAV0549	Serotonin 5HT1C re
36	65	12.2	514	20	AAV05896	Vicia sativa fatty
37	65	12.2	514	20	AAV05896	Drosophila melanog
38	64.5	12.1	59	22	AAAB6657	Human immune/haema
39	64.5	12.1	173	23	ABB72349	Murine protein iso
40	64.5	12.1	173	21	AAAG47017	Arabidopsis thalia
41	64.5	12.1	184	21	AAAG47016	Arabidopsis thalia
42	64.5	12.1	189	22	AAU27511	Human G-Protein Co
43	64.5	12.1	213	21	AAAG14923	Arabidopsis thalia
44	64.5	12.1	213	21	AAAG14923	Arabidopsis thalia
45	64.5	12.1	213	21	AAAG47015	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AAE13851
ID AAE13851 standard; peptide; 16 AA.

XX AAE13851;

XX 26-FEB-2002 (first entry)

DT Human T cell epitope related to lung tumour-specific protein.

DE Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;

KW antisense-therapy; vaccine; immune response; lung cancer;

KW T cell epitope.

XX Homo sapiens.

XX WO200172295-A2.

XX 04-OCT-2001.

XX 28-MAR-2001; 2001WO-US09991.

XX 29-MAR-2000; 2000US-0538037.

XX 05-JUN-2000; 2000US-0588937.

XX 18-AUG-2000; 2000US-0640878.

XX 22-SEP-2000; 2000US-234517P.

XX 01-NOV-2000; 2000US-0704512.

XX 14-DEC-2000; 2000US-0738973.

(CORI-) CORIXA CORP.

XX Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indrias CY;

PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;

the adjuvant relates to human polynucleotides (AAI79941-AAI93841) and the encoded proteins (AAO00010-AAO3910) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy vaccines or

the invention relates to human polypeptides (AAI7994I-AAI9384I) and the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or

us-09-854-133-586.rag

Wed Jul 9 15:44:27 2003

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 46 AA; 13.4%; Score 71.5; DB 22; Length 46;
 Query Match Best Local Similarity 81.0%; Pred. No. 0.69; Mismatches 1; Gaps 1;
 Matches 17; Conservative

QY 14 SETLSQTELEKKKKRERK 34
 DB 12 SETLSQKE-RKKERKKRERK 31

RESULT 4
 ID AAO09875 standard; Protein; 141 AA.

XX AC AAO09875;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 23767.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX PS WPI; 2001-514838/56.

XX DR N-PSDB; AAI89806.

XX PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -

XX PS Claim 20; SEQ ID NO 23767; 1399pp + Sequence Listing; English.

XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.

XX CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 141 AA;

Query Match 13.4%; Score 71.5; DB 22; Length 141;
 Best Local Similarity 31.0%; Pred. No. 2.5; Mismatches 9; Gaps 3;
 Matches 26; Conservative

QY 1 EVEVSRD--HASLGDSETLSQTELEKKKKRERK-----RKFOANGCIGDIFLFWIL 52
 DB 5 KVAVSVDLTMSNLSNETLSQKEKTKTKTKKQKTYEGRKKHKRCPTLLIREIQIT 64

QY 53 LFSHHWQI-ESLLCPPSPKVEVTCR 75

DB 65 TVRYHETHIRIVLLPQKQKNCWCR 88

RESULT 5
 ID AAG75462 standard; Protein; 163 AA.

XX AC AAG75462;

XX DT 03-SEP-2001 (first entry)

XX DE Human colon cancer antigen protein SEQ ID NO:6226.

XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma.

XX OS Homo sapiens.

XX PN WO200122920-A2.

XX PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26524.

XX PR 29-SEP-1999; 99US-0157137.

XX PR 03-NOV-1999; 99US-0163280.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX PS WPI; 2001-235357/24.

XX DR N-PSDB; AAI84867.

XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -

XX PS Claim 11; Page 7674-7676; 9803pp; English.

XX CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing P.
 CC Inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell,
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAH77789 represent sequences used in the exemplification of the
 CC present invention. 682 and page 7053 of the sequence listing were
 CC N.B. Pages 666 to 1052, 7921 and 7922.

XX SQ Sequence 163 AA;

Query Match 13.4%; Score 71.5; DB 22; Length 163;
 Best Local Similarity 38.6%; Pred. No. 2.9;

	Matches	22: Conservative	5: Mismatches	9: Indels	21: Gaps	4:
Oy	10	SLGD-SETLSQTELKRRKKRKFQANCIDFIWFILFSSHHTQESLLC				
			: : :			
db	11	SLGDKSTLSP----	RKKKKKKN-----	WIAW--LYGHSMAQAFC	47	

ABB03142
ID ABB03142 standard; Protein; 66 AA;
XX

08-JAN-2002 (first entry)

human musculoskeletal system related polypeptide SEQ ID NO 1089.

Homo sapiens.
WO200155367-A1.

02-AUG-2001.

17-JAN-2001;

31-JAN-2000; 2000US-0179065

24-FEB-2000; 2000US-0184664

16-MAR-2000; 2000US-01899874

18-APR-2000; 2000US-0198133

07-JUN-2000; 2000US-0209467.

30-JUN-2000; 2000US-0215135

07-JUL-2000; 2000US-0216880

11-JUL-2000; 2000US-0217496

26-JUL-2000; 2000US-0220963

14-AUG-2000; 2000US-0224518

14-AUG-2000; 2000US-0225213

14-AUG-2000; 2000US-0225266

14-AUG-2000; 2000US-0225268

14--AUG-2000; 2000US-0225447

14-AUG-2000; 2000US-0225758

18-AUG-2000; 2000US-0226279

22-AUG-2000; 2000US-0226868

23-AUG-2000; 2000US-0227009

01-SEP-2000; 2000US-0229287

01-SEP-2000; 2000US-0229344

03-SEP-2000; 2000US-0229509;

Abstract

PR	17-NOV-2000;	2000US-0249217.	
PR	17-NOV-2000;	2000US-0249218.	
PR	17-NOV-2000;	2000US-0249244.	
PR	17-NOV-2000;	2000US-0249245.	
PR	17-NOV-2000;	2000US-0249264.	
PR	17-NOV-2000;	2000US-0249265.	
PR	17-NOV-2000;	2000US-0249297.	
PR	17-NOV-2000;	2000US-0249299.	
PR	17-NOV-2000;	2000US-0249300.	
PR	01-DEC-2000;	2000US-0250160.	
PR	01-DEC-2000;	2000US-0250391.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	05-DEC-2000;	2000US-0256719.	
PR	06-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251969.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
XX			
XX	(HUMA-) HUMAN GENOME SCI INC.		
PA			
XX	Rosen CA, Barash SC, Ruben SM;		
PI			
XX	WPI; 2001-451937/48.		
DR	N-PSDB; AAL34724.		
DR			
XX	Isolated polypeptide for treating, preventing and/ or prognosing		
PT	disorders related to the musculoskeletal system including		
PT	musculoskeletal cancers and also for testing and detection e.g.		
PT	diagnosis		
XX			
XX	Claim 11; SEQ ID NO 1089; 781pp + Sequence Listing; English.		
PS			
XX	The invention relates to novel genes (AAL34669-AAL37666) and proteins		
XX	(ABB03087-ABB04109) associated with the musculoskeletal system useful		
CC	for preventing, treating or ameliorating medical conditions e.g. by		
CC	protein or gene therapy. The genes are isolated from a range of human		
CC	tissues disclosed in the specification. The nucleic acids, proteins,		
CC	antibodies and (ant)agonists are useful in the diagnosis, treatment		
CC	and prevention of: (a) cancer, e.g. breast and ovarian cancer and		
CC	other cancers of the adrenal gland, bone, bone marrow, breast,		
CC	gastrointestinal tract, liver, lung, or urogenital; (b) immune		
CC	disorders e.g. Addison's disease, allergies, autoimmune haemolytic		
CC	anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,		
CC	multiple sclerosis, rheumatoid arthritis and ulcerative colitis;		
CC	(c) cardiovascular disorders such as myocardial ischaemia; (d) wound		
CC	healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;		
CC	and (f) infectious diseases such as viral, bacterial, fungal and		
CC	parasitic infections.		
CC	Note: The sequence data for this patent did not form part of the		
CC	printed specification, but was obtained in electronic format directly		
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.		
XX			
XX	Sequence 66 AA;		
XX			
XX	Query Match 13.1%; Score 69.5; DB 22; Length 66;		
XX	Best Local Similarity 50.0%; Pred. No. 1.8;		
XX	Matches 19; Conservative 5; Mismatches 7; Indels 7; Gaps 2;		
QY	1 EVEVSRDHASL-----GDSEYLSQTELRKKKKRKKR 34		
DB	1 :		
	5 EAAVSCDHAGVLPQGHSDLSQ---NKKRKKRKKR 39		
	1 :		
RESULT 7			
ABBI7957			
ID	ABBI7957 standard; Protein; 63 AA.		

RESULT 8	
AAMS1507	
ID	AAMS1507 standard; Protein; 423 AA.
XX	
XX	
XX	AAMS1507;
XX	
XX	
DT	02-JAN-2002 (first entry)
DT	
DT	
DE	Mouse 5HT receptor.
XX	
XX	
KW	Mouse; 5HT receptor; chimera; cDNA library preparation; serotonin;
KW	G protein-coupled receptor.
XX	
OS	Mus musculus.
XX	
PN	JP2001178476-A.
XX	
PD	03-JUL-2001.

XX 28-DEC-1999; 99JP-0373989.
 PF XX
 PR 28-DEC-1999; 99JP-0373989.
 XX
 PA (AGENCY) AGENCY OF IND SCI & TECHNOLOGY.
 PA (KUBO/) KUBO Y.
 PA (KEIZ-) KEIZAI SANGYOSHO SANGYO GIJUTSU SOGO KEN.
 XX
 DR WPI; 2001-599907/68.
 DR N-PSDB; AAI69851.
 XX
 XX Preparing a chimera cDNA library comprises recombinantly expressing
 PT cDNA with a protein translation frame and a cDNA containing a carboxy
 PT terminal side from a stored region of a functional protein synthesized
 PT -
 XX
 PS Example 7; Fig 6; 24pp; Japanese.
 CC The present sequence is provided in a specification relating to the
 CC preparation of a chimera cDNA library containing a cDNA encoding a
 CC functional protein. The specification describes the preparation of
 CC a serotonin-related cDNA library, and the preparation of a
 CC G protein-coupled receptor chimera cDNA library from which an orphan
 CC G protein-coupled receptor cDNA was isolated.
 XX
 SQ Sequence 423 AA;
 Query Match 12.8%; Score 68; DB 22; Length 423;
 Best Local Similarity 22.8%; Pred. No. 22;
 Matches 29; Conservative 11; Mismatches 41; Indels 46; Gaps 5;
 QY 10 SLGSETLSQTELRKKRKKRKFQANGCIDFIIFWI---FWILL----- 53
 DB 271 TIGDVTTEQYLALR---RKKKTTVKMLVVLVLCWFFPNCYVLLSSKAIHTNNALY 327
 QY 54 FSHHWIQESLLC-----PPSPKEVTCREMLTGGLPWAT 87
 DB 328 FAFHWFAMSTCYNPFYICWLNENFRVELKALLSMCQRPKPQEDRLPSPVPSFVAVTE 387
 QY 88 RSHLGR 94
 DB 388 KSH-GRR 393
 QY
 DB
 RESULT 9
 AAE03740 standard; Protein; 423 AA.
 ID AAE03740
 XX
 AC AAE03740;
 XX
 DT 04-SEP-2001 (first entry)
 DE
 DE Mouse G-protein coupled receptor (GPCR) RP-23 protein.
 XX
 KW Mouse; G-protein coupled receptor; GPCR; RP-23; sphingosine-1-P04;
 KW sphingosine-1-phosphate; therapy; cardiovascular disease; cancer;
 KW inflammatory disease; cell signalling.
 XX
 OS Mus sp.
 XX
 XX WO200144439-A2.
 PN
 XX
 PD 21-JUN-2001.
 XX
 PF 15-DEC-2000; 2000WO-SE02563.
 XX
 PR 17-DEC-1999; 99SE-0004660.
 XX
 XX (ASTR) ASTRAZENECA AB.
 PA
 XX Ahmad S, Lembo P, Walker P;
 PI
 XX

DR WPI; 2001-398140/42.
 DR N-PSDB; AAD08133.
 XX
 PT Determining modulator of binding of sphingosine-1-P04 to RP-23
 PT receptor, involves incubating sphingosine-1-P04 and RP-23 receptor with
 PT test compound, determining the binding and comparing with a control -
 XX
 PS Disclosure; Fig 3; 25pp; English.
 XX
 CC The present sequence is mouse G-protein coupled receptor (GPCR)
 CC RP-23 protein. RP-23 is a ligand for sphingosine-1-phosphate
 CC (sphingosine-1-P04) which is formed by the breakdown of sphingolipids
 CC and is involved in cell signalling. Sphingosine-1-P04 causes an increase
 CC in receptor-generated adenylyl cyclase activity and an increase in
 CC intracellular calcium concentration. The present invention relates to
 CC methods for determining whether a test compound modulates interactions
 CC between sphingosine-1-P04 and RP-23 a particular G-protein coupled
 CC receptor. The method is useful for determining a modulator which binds
 CC sphingosine-1-P04 to an RP-23 receptor. The compounds identified are
 CC useful as therapeutic agents for treating cardiovascular diseases,
 CC inflammatory diseases and cancers.
 XX
 SQ Sequence 423 AA;
 Query Match 12.8%; Score 68; DB 22; Length 423;
 Best Local Similarity 22.8%; Pred. No. 22;
 Matches 29; Conservative 11; Mismatches 41; Indels 46; Gaps 5;
 QY 10 SLGSETLSQTELRKKRKKRKFQANGCIDFIIFWI---FWILL----- 53
 DB 271 TIGDVTTEQYLALR---RKKKTTVKMLVVLVLCWFFPNCYVLLSSKAIHTNNALY 327
 QY 54 FSHHWIQESLLC-----PPSPKEVTCREMLTGGLPWAT 87
 DB 328 FAFHWFAMSTCYNPFYICWLNENFRVELKALLSMCQRPKPQEDRLPSPVPSFVAVTE 387
 QY 88 RSHLGR 94
 DB 388 KSH-GRR 393
 QY
 DB
 RESULT 10
 AAE20598
 ID AAE20598 standard; Protein; 423 AA.
 XX
 AC AAE20598;
 XX
 DT 01-JUL-2002 (first entry)
 DE
 DE Homo sapiens glucocorticoid-induced receptor protein.
 XX
 KW Human; G-protein coupled receptor; GPCR; transgenic animal; receptor;
 KW genetic therapy; pharmacological; genetic disease; neuropsychological;
 KW neurological; psychotic illness; nephrotropic; gynaecological;
 KW psychostimulant; glucocorticoid-induced receptor.
 OS Homo sapiens.
 XX
 PN WO200203793-A2.
 XX
 PD 17-JAN-2002.
 XX
 PF 10-JUL-2001; 2001WO-US21923.
 XX
 PR 10-JUL-2000; 2000US-217058P.
 PR 10-JUL-2000; 2000US-217179P.
 PR 10-JUL-2000; 2000US-217223P.
 PR 10-JUL-2000; 2000US-217253P.
 PR 10-JUL-2000; 2000US-217255P.
 PR 10-JUL-2000; 2000US-217256P.
 PR 10-JUL-2000; 2000US-217257P.
 PR 11-JUL-2000; 2000US-217347P.
 PR 11-JUL-2000; 2000US-217629P.

PR 12-JUL-2000; 2000US-217537P.
 PR 12-JUL-2000; 2000US-218069P.
 PR 12-JUL-2000; 2000US-218074P.
 PR 27-JUL-2000; 2000US-218358P.
 PR 27-JUL-2000; 2000US-221483P.
 PR 07-AUG-2000; 2000US-223120P.
 PR 07-AUG-2000; 2000US-223122P.
 PR 26-OCT-2000; 2000US-243958P.
 PR 15-NOV-2000; 2000US-249408P.
 PR 20-NOV-2000; 2000US-252299P.
 PR 16-JAN-2001; 2001US-262113P.
 PR 16-JAN-2001; 2001US-262205P.
 XX (DELT-) DELTAGEN INC.

XX Allen KD, Brennan TJ;

XX WPI; 2002-164574/21.

XX N-PSDB; RAD32919.

PT Novel non-human transgenic animal, especially transgenic mice useful
 PT for identifying an agent that modulates expression or function of
 PT target gene, comprises disruptions in target G protein-coupled receptor
 PT gene.

XX Disclosure; Fig 7; 101pp; English.

XX The invention relates to a non-human transgenic animal having targeted
 CC G-protein coupled receptor (GPCR) gene disruptions in melanocortin-3 gene
 CC 5-HT-2B gene, chemokine receptor 9A gene, glucocorticoid-induced receptor
 CC gene, orphan GPR10 (UHR)-1 gene, orphan GPR14 gene, orphan GPR15 gene,
 CC beta chemokine receptor (EGL) gene, endothelial differentiation GPCR3
 CC (EDG3) gene, ATP receptor P2U1 gene or adenosine 3 receptor gene. The
 CC transgenic animal is useful for identifying an agent that modulates the
 CC expression or function of the target gene, for identifying an agent that
 CC ameliorates a phenotype associated with a disruption in the target gene.
 CC A transgenic construct is useful for producing a transgenic animal,
 CC preferably a transgenic mouse. The transgenic animal is useful for
 CC testing the efficacy of proposed genetic and pharmacological therapies
 CC for human genetic diseases, such as neurological, neuropsychological or
 CC psychotic illnesses. The transgenic animal is also useful as models for
 CC diseases, disorders or conditions associated with phenotypes relating to
 CC a disruption in a target, and to identify pharmaceuticals, therapies,
 CC drugs and interventions which may be effective in treating a disease or
 CC other phenotypic characteristics of the animal. An agent which modulates
 CC the expression of the target gene is useful as a therapeutic for treating
 CC conditions associated with a disruption of the target gene. The present
 CC sequence is human glucocorticoid-induced receptor protein.

XX Sequence 423 AA;

Query Match 12.8%; Score 68; DB 23; Length 423;
 Best Local Similarity 22.8%; Pred. No. 22;
 Matches 29; Conservative 11; Mismatches 41; Indels 46; Gaps 5;

QY 10 SLGDSLTSLQELKRRKKRRKFOANGGIDFIIFIWI---FWILL----- 53
 Db 271 TIGDVTTEQYIALR---RRKRTTVKMLVLVLFALCWFPLNCLVSSKAIHTNNALY 327

QY 54 FSHHWIOESLLC-----PPSPKVEVTCREMLTGCLPWAT 87
 Db 328 FAFHWFAMSTCYNPFIYCNLENFRVELKALLSMCQRPKPQEDRFLPSVPSFVATE 387

QY 88 RSHLGR 94
 Db 388 KSH-GRR 393

RESULT 11

ID AAY28851

XX AAY28851 standard; Protein; 458 AA.

AC AAY28851;

XX

DT

XX

DE

XX

KW

KW

KW

KW

KW

KW

OS

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

25-JAN-2000 (first entry)

Active form of human 5-HT2C serotonin receptor (AP-1).

Serotonin receptor; 5-hydroxytryptamine; human 5-HT2C; non-endogenous;
 active form; point mutation; third intracellular loop; specificity;
 linkage; transduction pathway; biological response; antagonist; AP-1;
 baseline intracellular response; compound; treatment; prevention; agent;
 central nervous system; cardiovascular disorder; drug screening;
 gastric or inflammatory disorder; constitutive expression.

Homo sapiens.

Synthetic.

Key Location/Qualifiers

Misc-difference 310

/note= "Ser in endogenous form of 5-HT2C replaced by Lys"

W09952927-A1.

21-OCT-1999.

14-APR-1999; 99WO-US08168.

14-APR-1998; 98US-0060188.

26-JUN-1998; 98US-0090783.

18-DEC-1998; 98US-0112909.

05-MAR-1999; 98US-0123000.

(AREN-) ARENA PHARM INC.

(TRIP-) TRIPOS INC.

Behan DP, Chalmers DT, Foster RJ, Glen RC, Lawless MS, Liaw CW;

Liu Q, Russo JF, Smith JR, Thomsen WJ;

WPI; 1999-611285/52.

N-PSDB; AA208098.

Novel nucleic acid, used to identify inverse agonists e.g. for treating

or preventing cardiovascular disease

Claim 2; Fig 5b; 141pp; English.

The present protein sequence is the non-endogenous, constitutively active

form of human 5-HT2C serotonin receptor. A point mutation replaces a Ser

with Lys in the third intracellular loop of the 5-HT2C receptor. This

activated receptor allows linkage to the transduction pathway and

produces a biological response. This sequence can be used to identify

antagonists that bind serotonin receptors and reduces the baseline

intracellular response caused by receptor activation. Constitutive

expression of this receptor is suitable for drug screening and to

identify agents of very high specificity. These compounds are potentially

useful for treating or preventing disorders involving central nervous

system and cardiovascular, gastric or inflammatory disorders, that

involves these receptors.

Sequence 458 AA;

Query Match 12.8%; Score 68; DB 20; Length 458;

Best Local Similarity 27.4%; Pred. No. 24;

Matches 23; Conservative 15; Mismatches 28; Indels 18; Gaps 3;

QY 6 RDHSLGDSLTSLQELKRRKKRRK-----ERRFOANGGIDFIIFIWIFWILLF 54

Db 269 RNTAEENSANPNQONARRKKRRPRCTMQAINNERKAKKVLGIVFFVFLIMWCPPF 328

QY 55 SHHWIOESLLC-----PPSPKVEVTCREML 78

Db 329 ITNILL--SVLC-----EKSCNOKL 345

RESULT 12

XX AC AAB93643;
 XX DT 26-JUN-2001 (first entry)
 XX DE Human protein sequence SEQ ID NO:13140.
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX OS Homo sapiens.
 XX PN EP1074617-A2.
 XX PD 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-0116126.
 XX PR 29-JUL-1999; 99JP-0248036.
 XX PR 27-AUG-1999; 99JP-0300253.
 XX PR 11-JAN-2000; 2000JP-0118776.
 XX PR 02-MAY-2000; 2000JP-0183767.
 XX PR 09-JUN-2000; 2000JP-0241899.
 XX PA (HELI-) HELIX RES INST.
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX DR WPI; 2001-318749/34.
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
 XX PT full-length cDNAs defined in the specification, and for the detection
 XX PT and/or diagnosis of the abnormality of the proteins encoded by the
 XX PT full-length cDNAs -
 XX PS Claim 8; SEQ ID 13140; 2537pp + CD ROM; English.
 XX CC The present invention describes primer sets for synthesizing 5602
 XX CC full-length cDNAs defined in the specification. Where a primer set
 XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 XX CC to the complementary strand of a polynucleotide which comprises one of
 XX CC the 5602 nucleotide sequences defined in the specification, where the
 XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 XX CC of an oligonucleotide comprising a sequence complementary to the
 XX CC complementary strand of a polynucleotide which comprises a 5'-end
 XX CC sequence and an oligonucleotide comprising a sequence complementary to a
 XX CC polynucleotide which comprises a 3'-end sequence, where the
 XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
 XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
 XX CC the specification. The primer sets can be used in antisense therapy and
 XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 XX CC particularly full-length cDNAs. The primers are also useful for the
 XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
 XX CC the full-length cDNAs. The primers allow obtaining of the full-length
 XX CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 XX CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 XX CC represent oligonucleotides, all of which are used in the exemplification
 XX CC of the present invention.
 XX SQ Sequence 105 AA;
 Query Match 12.7%; Score 67.5; DB 22; Length 105;
 Best Local Similarity 41.7%; Pred. No. 5.1;
 Matches 15; Conservative 9; Mismatches 7; Indels 5; Gaps 1;
 QY 3 EVSRDHALGDSSETLSQTELRKKKKRERKFOAN 38
 Db 43 EIAPLHSLGDS-----TLRLKKKKKKKKRKRN 73
 RESULT 15
 AABP00516

ID ABP00516 standard; Protein; 81 AA.
 XX AC AABP00516;
 XX DT 24-JUN-2002 (first entry)
 XX DE Human ORFX protein sequence SEQ ID NO:1014.
 XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.
 XX OS Homo sapiens.
 XX PN WO200192523-A2.
 XX PD 06-DEC-2001.
 XX PF 29-MAY-2001; 2001WO-US10836.
 XX PR 30-MAY-2000; 2000US-206132P.
 XX PR 29-AUG-2000; 2000US-228716P.
 XX PA (CURA-) CURAGEN CORP.
 XX PI Shinkets RA, Leach MD;
 XX DR WPI; 2002-106308/14.
 XX DR N-PSDB; ABN16268.
 XX PT Novel human polypeptides and polynucleotides useful for diagnosing,
 XX PT preventing and treating cardiovascular disease, neurodegenerative,
 XX PT hyperproliferative disorders and autoimmune disorders -
 XX PS Disclosure; SEQ ID 1014; 1037pp; English.
 XX CC The present invention describes substantially purified human proteins
 XX CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 XX CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 XX CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 XX CC treating or preventing a pathology associated with an ORFX-associated
 XX CC disorder in humans, and in the manufacture of a medicament for treating a
 XX CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 XX CC sequences can be used in gene therapy. ORFX sequences can be used in the
 XX CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 XX CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 XX CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 XX CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 XX CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 XX CC storage disease, various immune deficiencies and disorders, infectious
 XX CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 XX CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 XX CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 XX CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 XX CC bone degenerative disorders, or periodontal disease, and for gut
 XX CC protection or regeneration and treatment of lung or liver fibrosis,
 XX CC reperfusion injury in various tissues and conditions resulting from
 XX CC systemic cytokine damage.
 XX CC N.B. The sequence data for this patent did not form part of the printed
 XX CC specification, but was obtained in electronic format directly from WIPO
 XX CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 81 AA;
 Query Match 12.6%; Score 67; DB 23; Length 81;
 Best Local Similarity 48.3%; Pred. No. 4.3;
 Matches 14; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

OY 18 SOTELRKKRKKRERKFOANCGIDFIIF 46
| :|||| :|||| : | ||
Db 11 SHRQLRKKAKKKRERARINTFCVFIFFIF 39

Search completed: July 9, 2003, 15:07:40
Job time : 56.0796 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	68	12.8	458	3	US-09-292-071-29		Sequence 29, Appl
2	68	12.8	458	4	US-09-292-069A-29		Sequence 29, Appl
3	68	12.8	458	4	US-09-767-013-29		Sequence 29, Appl
4	67	12.6	458	1	US-08-310-271-2		Sequence 27, Appl
5	67	12.6	458	3	US-09-292-071-27		Sequence 27, Appl
6	67	12.6	458	4	US-09-292-069A-27		Sequence 27, Appl
7	67	12.6	458	4	US-09-032-742-10		Sequence 10, Appl
8	67	12.6	458	4	US-09-767-013-27		Sequence 27, Appl
9	66	12.4	478	3	US-09-292-071-33		Sequence 33, Appl
10	66	12.4	478	4	US-09-292-069A-33		Sequence 33, Appl
11	66	12.4	478	4	US-09-767-013-33		Sequence 33, Appl
12	64	12.0	380	4	US-09-134-001C-3830		Sequence 3830, Ap
13	64	12.0	488	4	US-09-215-694-15		Sequence 15, Appl
14	62	11.7	478	3	US-09-292-071-31		Sequence 31, Appl
15	62	11.7	478	4	US-09-292-069A-31		Sequence 31, Appl
16	62	11.7	478	4	US-09-767-013-31		Sequence 31, Appl
17	61.5	11.6	111	1	US-08-334-773A-6		Sequence 6, Appl
18	61.5	11.6	365	2	US-08-428-243-9		Sequence 9, Appl
19	61.5	11.6	365	5	PCF-US93-10301-9		Sequence 9, Appl
20	61.5	11.6	422	1	US-07-996-772A-12		Sequence 12, Appl
21	61.5	11.6	445	2	US-08-157-185-2		Sequence 2, Appl
22	61.5	11.6	445	3	US-08-281-926B-2		Sequence 2, Appl
23	61.5	11.6	445	4	US-09-450-797-2		Sequence 2, Appl
24	61.5	11.6	445	4	US-09-328-314-16		Sequence 16, Appl
25	61.5	11.6	445	4	US-09-450-790A-2		Sequence 2, Appl
26	61.5	11.6	445	4	US-09-332-837-2		Sequence 2, Appl
27	61.5	11.6	445	5	PCF-US93-10553-2		Sequence 2, Appl

100

```

Matches      23; Conservative      14; Mismatches      29; Indels      18; Gaps      3;

y          6 RDHASLGDSETLSQTELKRRKKR-----ERKFQANCNGIDFIIFWIFWILLF 54
           | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
b          269 RNTAAEENSANPDQNQARRKKRRPRGTMQAINNERRKASKVLGIVFFVFLIMWCPPF 328
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

y          55 SHHWIOESLLCPPSPKEVTCREML 78
           : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
b          329 ITNIL--SVLC-----EKSCNQKL 345

RESULT 5
US-09-292-071-27
Sequence 27, Application US/09292071
Patent No. 6107324
GENERAL INFORMATION:
APPLICANT: Behan, Dominic
APPlicant: Chalmers, Derrick
TITLE OF INVENTION: NO. 6107324-Endogenous, Constitutively Activated
TITLE OF INVENTION: Human Serotonin Receptors and Small Molecule Modulators Thereof
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arena Pharmaceuticals, Inc.
STREET: 6166 Nancy Ridge Drive
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/292,071
FILING DATE: April 14, 1999
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Rosen
REGISTRATION NUMBER: 39,822
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 564-6525
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 458 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-09-292-071-27

Query Match      12.6%; Score 67; DB 3; Length 458;
Best Local Similarity 27.4%; Pred. No. 7.2;
Matches      23; Conservative      14; Mismatches      29; Indels      18; Gaps      3;

QY          6 RDHASLGDSETLSQTELKRRKKR-----ERKFQANCNGIDFIIFWIFWILLF 54
           | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db          269 RNTAAEENSANPDQNQARRKKRRPRGTMQAINNERRKASKVLGIVFFVFLIMWCPPF 328
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY          55 SHHWIOESLLCPPSPKEVTCREML 78
           : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db          329 ITNIL--SVLC-----EKSCNQKL 345

RESULT 6
US-09-292-069A-27
Sequence 27, Application US/09292069A
Patent No. 6140509
GENERAL INFORMATION:
APPLICANT: Behan, Dominic P
APPlicant: Chalmers, Derek T

```

REGISTRATION NUMBER: 27,965
REFERENCE/DOCKET NUMBER: 3086-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 431-1703
TELEFAX: (610) 431-4181
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 458 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-09-032-742-10

Query Match 12.6%; Score 67; DB 4; Length 458;
Best Local Similarity 27.4%; Pred. No. 7.2;
Matches 23; Conservative 14; Mismatches 29; Indels 18; Gaps 3;

QY 6 RDHASLGDSETLSQTELRKKRKKR-----ERKFOANGCIDFIIFWIFWILLF 54
Db 269 RNTAEENSANPDQNNARRKKRRPRGTMQAINNERKASKVLGIVFFVFLIMWCPFF 328
QY 55 SHHWIQESLLCPPSPKVTCTREML 78
Db 329 ITNIL--SVLC-----EKSCNQKL 345

RESULT 8

US-09-767-013-27
Sequence 27, Application US/09767013
Patent No. 6420541
GENERAL INFORMATION:
APPLICANT: Behan, Dominic

APPLICANT: Chalmers, Derek T
TITLE OF INVENTION: No. 6420541-Endogenous, Constitutively Activated Human
TITLE OF INVENTION: Serotonin Receptors and Small Molecule Modulators
FILE REFERENCE: AREN0035
CURRENT APPLICATION NUMBER: US/09/767,013
CURRENT FILING DATE: 2001-01-23
PRIOR APPLICATION NUMBER: 09/292,072
PRIOR FILING DATE: 1999-04-14
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 27
LENGTH: 458
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: No. 6420541el Sequence

US-09-767-013-27

Query Match 12.6%; Score 67; DB 4; Length 458;
Best Local Similarity 27.4%; Pred. No. 7.2;
Matches 23; Conservative 14; Mismatches 29; Indels 18; Gaps 3;

QY 6 RDHASLGDSETLSQTELRKKRKKR-----ERKFOANGCIDFIIFWIFWILLF 54
Db 269 RNTAEENSANPDQNNARRKKRRPRGTMQAINNERKASKVLGIVFFVFLIMWCPFF 328
QY 55 SHHWIQESLLCPPSPKVTCTREML 78
Db 329 ITNIL--SVLC-----EKSCNQKL 345

RESULT 9

US-09-292-071-33
Sequence 33, Application US/09292071
Patent No. 6107324
GENERAL INFORMATION:
APPLICANT: Behan, Dominic

APPLICANT: Chalmers, Derek T
TITLE OF INVENTION: No. 6107324-Endogenous, Constitutively Activated Human
TITLE OF INVENTION: Serotonin Receptors And Small Molecule Modulators
FILE REFERENCE: AREN0033
CURRENT APPLICATION NUMBER: US/09/292,069A
CURRENT FILING DATE: 1999-04-14
PRIOR APPLICATION NUMBER: 60/090,783
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/112,909
PRIOR FILING DATE: 1998-12-18
PRIOR APPLICATION NUMBER: 60/123,000
PRIOR FILING DATE: 1999-03-05

TITLE OF INVENTION: Human Serotonin Receptors and Small Molecule Modulators
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arena Pharmaceuticals, Inc.
STREET: 6166 Nancy Ridge Drive
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121

COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/292,071
FILING DATE: April 14, 1999

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Rosen
REGISTRATION NUMBER: 39,822
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 564-6525
TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 478 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-09-292-071-33

Query Match 12.4%; Score 66; DB 3; Length 478;
Best Local Similarity 25.8%; Pred. No. 9.8;
Matches 23; Conservative 16; Mismatches 32; Indels 18; Gaps 3;

QY 6 RDHASLGDSETLSQTELRKKRKKR-----ERKFOANGCIDFIIFWIFWILLF 54
Db 289 RNTAEENSANPDQNNARRKKRRPRGTMQAINNERKASKVLGIVFFVFLIMWCPFF 348
QY 55 SHHWIQESLLCPPSPKVTCTREMLTGCL 83
Db 349 ITNIM--AVICKES-----CNEDVIGALL 370

RESULT 10

US-09-292-069A-33

Sequence 33, Application US/09292069A

Patent No. 6140509

GENERAL INFORMATION:

APPLICANT: Behan, Dominic P

APPLICANT: Chalmers, Derek T

APPLICANT: Foster, Richard J

APPLICANT: Glen, Robert C

APPLICANT: Lawless, Michael S

APPLICANT: Liaw, Chen W

APPLICANT: Liu, Qian

APPLICANT: Russo, Joseph F

APPLICANT: Smith, Julian R

APPLICANT: Thomsen, William J

TITLE OF INVENTION: No. 6140509-Endogenous, Constitutively Activated Human

TITLE OF INVENTION: Serotonin Receptors And Small Molecule Modulators

FILE REFERENCE: AREN0033

CURRENT APPLICATION NUMBER: US/09/292,069A

CURRENT FILING DATE: 1999-04-14

PRIOR APPLICATION NUMBER: 60/090,783

PRIOR FILING DATE: 1998-06-26

PRIOR APPLICATION NUMBER: 60/112,909

PRIOR FILING DATE: 1998-12-18

PRIOR APPLICATION NUMBER: 60/123,000

PRIOR FILING DATE: 1999-03-05


```

/ TITLE OF INVENTION: NO. 6107324-Endogenous, Constitutively Activated
/
/ TITLE OF INVENTION: Human Serotonin Receptors and Small Molecule Modulators
/
/ NUMBER OF SEQUENCES: 33
/
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Arena Pharmaceuticals, Inc.
/ STREET: 6166 Nancy Ridge Drive
/ CITY: San Diego
/ STATE: CA
/ COUNTRY: USA
/
/ ZIP: 92121
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk

```


Sequence 673, App
Sequence 24, Appl
Sequence 155, App
Sequence 38925, A
Sequence 37546, A
Sequence 41, Appl
Sequence 33, Appl
Sequence 32, Appl
Sequence 16, Appl
Sequence 40, Appl
Sequence 71, Appl
Sequence 4, Appl
Sequence 3, Appl
Sequence 43, Appl
Sequence 1124, Ap
Sequence 1, Appl
Sequence 3, Appl
Sequence 5, Appl
Sequence 7, Appl
Sequence 9, Appl
Sequence 11, Appl
Sequence 13, Appl
Sequence 17, Appl
Sequence 19, Appl
Sequence 21, Appl
Sequence 23, Appl

173 9 US-09-866-050A-673
487 10 US-09-801-368-24
53 9 US-10-001-857-155
150 10 US-09-864-761-38925
67 10 US-09-864-761-37546
432 9 US-09-877-843-41
445 9 US-09-877-843-39
445 9 US-10-225-567A-22
445 10 US-09-989-861-16
479 9 US-09-877-843-40
62 9 US-09-883-343A-71
61 11.5 641 9 US-10-242-343A-4
309 9 US-10-060-795B-3
448 9 US-09-877-843-43
451 10 US-09-764-864-1124
459 9 US-10-280-858-1
459 9 US-10-280-858-3
459 9 US-10-280-858-5
459 9 US-10-280-858-7
459 9 US-10-280-858-9
459 9 US-10-280-858-11
459 9 US-10-280-858-13
459 9 US-10-280-858-17
459 9 US-10-280-858-19
459 9 US-10-280-858-21
459 9 US-10-280-858-23

ALIGNMENTS

RESULT 1
US-09-854-133-586
; Sequence 586, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh A.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 586
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-133-586

Query Match 100.0%; Score 532; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. le-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EVEVSRDHASLGDSSETLSQTLERKKERKKERKQFQANCIDFIIFWIFWILLFHHWQ 60
Db 1 EVEVSRDHASLGDSSETLSQTLERKKERKKERKQFQANCIDFIIFWIFWILLFHHWQ 60
Qy 61 ELLCPPSPKVTCTREMLTGGCLPWATRSHLGRKCS 97
Db 61 ELLCPPSPKVTCTREMLTGGCLPWATRSHLGRKCS 97
RESULT 2
US-10-144-649A-586
; Sequence 586, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: July 9, 2003, 15:06:38 ; Search time 32.6195 Seconds
(without alignments)
346.196 Million cell updates/sec

Title: US-09-854-133-586
Perfect score: 532
Sequence: 1 EVEVSRDHASLGDSSETLSQTLERKKERKKERKQFQANCIDFIIFWIFWILLFHHWQ 97

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 445758 seqs, 116419773 residues 445758

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pcp.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
 - 5: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pcp.*
 - 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pcp.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pcp.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pcp.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	532	100.0	97	9	US-09-854-133-586
2	532	100.0	97	9	US-10-144-649A-586
3	532	100.0	97	10	US-09-738-973-586
4	532	100.0	114	9	US-10-144-649A-742
5	98	18.4	16	9	US-09-854-133-587
6	98	18.4	16	10	US-10-144-649A-587
7	98	18.4	16	10	US-09-738-973-587
8	80.5	15.1	119	9	US-09-892-877-400
9	80.5	15.1	119	9	US-09-948-783-400
10	71.5	13.4	163	9	US-10-106-698-6236
11	69.5	13.1	66	10	US-09-764-877-1089
12	68	12.8	423	9	US-09-903-396A-2
13	68	12.8	458	9	US-10-251-385-230
14	67	12.6	458	9	US-10-251-385-126
15	67	12.6	458	9	US-10-225-567A-16
16	67	12.6	1601	10	US-09-862-027-40
17	66.5	12.5	64	9	US-09-764-891-4290
18	66	12.4	369	10	US-09-838-955-3
19	66	12.4	673	9	US-10-157-031-291

APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Algate, Paul A.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
THE THERAPY AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.475C11
CURRENT APPLICATION NUMBER: US/10/144,649A
CURRENT FILING DATE: 2002-08-21
NUMBER OF SEQ ID NOS: 749
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 586
LENGTH: 97
TYPE: PRT
ORGANISM: Homo sapiens
US-10-144-649A-586

Query Match 100.0%; Score 532; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. 1e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EVEVSRDHASLGDSSETLSQTELKRRKKRERKFOANCIDFIIFWIFWILLFHHWQ 60
Db 1 EVEVSRDHASLGDSSETLSQTELKRRKKRERKFOANCIDFIIFWIFWILLFHHWQ 60
QY 61 ESLCPSPPKEVTCREMLTGGCLPWATRSHLGRKCS 97
Db 61 ESLCPSPPKEVTCREMLTGGCLPWATRSHLGRKCS 97

RESULT 3
US-09-738-973-586
Sequence 586, Application US/09738973
Patent No. US20020110563A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Henderson, Robert A.
APPLICANT: Lodes, Michael J.
APPLICANT: Fling, Steven P.
APPLICANT: Mohamath, Raodoh
APPLICANT: Algate, Paul A.
APPLICANT: Secrist, Heather
APPLICANT: Indirias, Carol Yoseph
APPLICANT: Benson, Darin R.
APPLICANT: Elliot, Mark
APPLICANT: Mannion, Jane
APPLICANT: Kalos, Michael D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
THE THERAPY AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.475C9
CURRENT APPLICATION NUMBER: US/09/738,973
CURRENT FILING DATE: 2000-12-14
NUMBER OF SEQ ID NOS: 587
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 586
LENGTH: 97
TYPE: PRT
ORGANISM: Homo sapiens
US-09-738-973-586

Query Match 100.0%; Score 532; DB 10; Length 97;
Best Local Similarity 100.0%; Pred. No. 1e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EVEVSRDHASLGDSSETLSQTELKRRKKRERKFOANCIDFIIFWIFWILLFHHWQ 60
Db 1 EVEVSRDHASLGDSSETLSQTELKRRKKRERKFOANCIDFIIFWIFWILLFHHWQ 60
QY 61 ESLCPSPPKEVTCREMLTGGCLPWATRSHLGRKCS 97
Db 61 ESLCPSPPKEVTCREMLTGGCLPWATRSHLGRKCS 97

RESULT 4
US-10-144-649A-742
Sequence 742, Application US/10144649A
Publication No. US20030118599A1
GENERAL INFORMATION:
APPLICANT: Lodes, Michael J.
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Algate, Paul A.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
THE THERAPY AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.475C11
CURRENT APPLICATION NUMBER: US/10/144,649A
CURRENT FILING DATE: 2002-08-21
NUMBER OF SEQ ID NOS: 749
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 742
LENGTH: 114
TYPE: PRT
ORGANISM: Homo sapiens
US-10-144-649A-742

Query Match 100.0%; Score 532; DB 9; Length 114;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EVEVSRDHASLGDSSETLSQTELKRRKKRERKFOANCIDFIIFWIFWILLFHHWQ 60
Db 18 EVEVSRDHASLGDSSETLSQTELKRRKKRERKFOANCIDFIIFWIFWILLFHHWQ 77
QY 61 ESLCPSPPKEVTCREMLTGGCLPWATRSHLGRKCS 97
Db 78 ESLCPSPPKEVTCREMLTGGCLPWATRSHLGRKCS 114

RESULT 5
US-09-854-133-587
Sequence 587, Application US/09854133
Publication No. US20020183499A1
GENERAL INFORMATION:
APPLICANT: Lodes, Michael J.
APPLICANT: Mohamath, Raodoh
APPLICANT: Henderson, Robert A.
APPLICANT: Benson, Darin R.
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
THE THERAPY AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.475C10
CURRENT APPLICATION NUMBER: US/09/854,133
CURRENT FILING DATE: 2001-05-11
NUMBER OF SEQ ID NOS: 735
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 587
LENGTH: 16
TYPE: PRT
ORGANISM: Homo sapiens
US-09-854-133-587

Query Match 18.4%; Score 98; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00064;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 FOANCIDFIIFWIFW 50
Db 1 FOANCIDFIIFWIFW 16

RESULT 6
US-10-144-649A-587
Sequence 587, Application US/10144649A
Publication No. US20030118599A1
GENERAL INFORMATION:

THE

```

1  GENERAL INFORMATION:
2  APPLICANT: LifeSpan Biosciences
3  APPLICANT: Brown, Joseph P.
4  APPLICANT: Burmer, Glenna C.
5  APPLICANT: Roush, Christine L.
6  TITLE OF INVENTION: ANTIGENIC PEPTIDES AND
7  FILE REFERENCE: 1920-4-4
8  CURRENT APPLICATION NUMBER: US/10/225,567A
9  CURRENT FILING DATE: 2001-12-19
10 PRIOR APPLICATION NUMBER: 60/257,144
11 PRIOR FILING DATE: 2000-12-19
12 NUMBER OF SEQ ID NOS: 2292
13 SOFTWARE: PatentIn version 3.1
14 SEQ ID NO 16
15 LENGTH: 458
16 TYPE: PRT
17 ORGANISM: Homo sapiens
18 US-10-225-567A-16

```

Query Match	12.6%	Score 67;	DB 9;	Length 458;
Best Local Similarity	27.4%;	Pred. No. 40;		
Matches	23;	Conservative 14;	Mismatches 29;	Indels 18;
				Gaps 3;
6	RDHSLGDSSETLSQTELKRRKKR	-----	ERKFOANCIGDIFIFWIFWILLF	54
	: : : : : : :			
269	RNTAEENSANPDQNAARRKKRRPRGTMQA	-----	INNERKASKVLGIVFVFLIMMCPFF	328
55	SHHWIQESLLCPPSPKEVTCREML			78
	: : : : : :			
329	ITNII--SVLC-----		EKSCNQKL	345

Search completed: July 9, 2003, 15:10:28
Job time : 33.6195 secs

APPLICANT: Behan, Dominic P.
 APPLICANT: Chalmers, Derek T.
 APPLICANT: Liaw, Chen W
 TITLE OF INVENTION: NO. US20030105292A1 - Endogenous, Constitutively Activated Human G
 TITLE OF INVENTION: Protein-Coupled
 TITLE OF INVENTION: ReceptoPloys
 FILE REFERENCE: AREN-0040
 CURRENT APPLICATION NUMBER: US/10/251,385
 CURRENT FILING DATE: 2002-09-20
 PRIOR APPLICATION NUMBER: US/09/170,496
 PRIOR FILING DATE: 1998-10-13
 NUMBER OF SEQ ID NOS: 294
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 230
 LENGTH: 458
 TYPE: PRT
 ORGANISM: Homo sapiens
 S-10-251-385-230

	Query Match	12.8%;	Score 68;	DB 9;	Length 458;
	Best Local Similarity	27.4%;	Pred. No. 31;		
	Matches	23;	Conservative	15;	Mismatches 28; Indels 18; Gaps 3;
6	RDHSLGDSLTLSQTELKRRKKRR-----ERKFAQNCIDITLFIWFIWLLF	54			
269	RNTAEENSANPNQDANRRKKRRPRGTMOAINNERAKKVLGIVFVFLIMCQPF	328			
55	SHHWIOESLCLPPSPKEVTCREML	78			
329	TNITL--SVLC-----EKSCNKL	345			

RESULT 14
 ; US-10-251-385-126
 ; Sequence 126, Application US/10251385
 ; Publication No. US20030105292A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Behan, Dominic P.
 ; APPLICANT: Chalmers, Derek T.
 ; APPLICANT: Liaw, Chen W.
 ; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
 ; TITLE OF INVENTION: Protein-Coupled
 ; TITLE OF INVENTION: Receptors
 ; FILE REFERENCE: AREN-0040
 ; CURRENT APPLICATION NUMBER: US/10/251,385
 ; CURRENT FILING DATE: 2002-09-20
 ; PRIOR APPLICATION NUMBER: US/09/170,496
 ; PRIOR FILING DATE: 1998-10-13
 ; NUMBER OF SEQ ID NOS: 294
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 126
 ; LENGTH: 458
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-251-385-126

Query Match	12.6%	Score 67;	DB 9;	Length 458;
Best Local Similarity	27.4%;	Pred. No. 40;		
Matches 23;	Conservative 14;	Mismatches 29;	Indels 18;	Gaps 3;
QY	6	RDHASLGDSSETLSQTELRRKKRKKR-----	ERRQFQACGIDGFIWFIFWILLF	54
		: : : : : : : : :		
Db	269	RTAAEENSANPDQNAARRKKERRRPGTMOA	INNERKASKVLGIVFFVFLIMCFF	328
QY	55	SHHWIQESLLCPSPKEVTCREML	78	
		: : : : : :		
Db	329	ITNIL--SVLC-----	EKSCNQKL	345

RESULT 15
US-10-225-567A-16
; Sequence 16, Application US/10225567A
; Publication No. US20030113798A1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 43.7788 Seconds
(without alignments)
213.004 Million cell updates/sec

Title: US-09-854-133-586
Perfect score: 532
Sequence: 1 EVEVSRDHASLGDSSETLSQT.....LTGGLPWATNRSHLGRKCS 97

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

pred, No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	72.5	13.6	177	A82037	ATP synthase Fl, d
2	70	13.2	684	T32024	hypothetical prote
3	68	12.8	381	A40470	glucocorticoid-ind
4	68	12.8	423	B40470	glucocorticoid-ind
5	68	12.8	443	D40470	glucocorticoid-ind
6	68	12.8	491	C40470	glucocorticoid-ind
7	67	12.6	458	JS0616	serotonin receptor
8	67	12.6	1851	T19864	hypothetical prote
9	66.5	12.5	2197	T19600	variant-specific s
10	66	12.4	132	T49635	hypothetical prote
11	66	12.4	673	F40201	artifact-warning s
12	65.5	12.3	207	A64915	yody protein homol
13	65.5	12.3	207	A90916	probable oxidoredu
14	65.5	12.3	207	F85764	probable oxidoredu
15	65	12.2	514	T08014	Cytochrome P450 CY
16	64.5	12.1	163	A48770	Ca2+-transporting
17	64	12.0	199	G89530	protein H28G03.3 [
18	62.5	11.7	177	S06079	H+-transporting tw
19	62.5	11.7	566	AB3061	30S ribosomal prot
20	62.5	11.7	572	F98225	probable bacteriop
21	62.5	11.7	728	B83131	variant-specific s
22	62.5	11.7	2228	T14029	hypothetical prote
23	62	11.7	326	T27655	hypothetical prote
24	62	11.7	350	T25366	hypothetical prote
25	62	11.7	356	T19520	hypothetical prote
26	62	11.7	428	T49196	nematode resistanc
27	62	11.7	1549	A40691	trichohyalin - she
28	62	11.7	1555	T18688	hypothetical prote
29	62	11.7	1973	G89608	protein B0272.5 [1

30	62	11.7	1973	2	T18686	hypothetical prote
31	62	11.7	2305	2	T15571	hypothetical prote
32	62	11.7	2305	2	B89608	protein C23F12.1 I
33	61.5	11.6	319	2	T23329	hypothetical prote
34	61.5	11.6	445	2	A48881	serotonin receptor
35	61.5	11.6	2164	1	GNNY89	genome polyprotein
36	61.5	11.6	2573	2	D71614	hypothetical prote
37	61.5	11.6	3026	2	T28431	variant surface pr
38	61	11.5	62	1	A48941	pediocin PA-1 prec
39	61	11.5	670	2	T51275	hypothetical prote
40	61	11.5	686	2	T25987	hypothetical prote
41	60.5	11.4	370	2	T48578	transcription fact
42	60.5	11.4	468	1	TVMSE2	ribosomal protein
43	60.5	11.4	568	1	R32R1	hypothetical prote
44	60.5	11.4	777	2	T21048	E4 protein - human
45	60	11.3	108	1	W4WL11	

ALIGNMENTS

RESULT 1

A82037
ATP synthase Fl, delta subunit VC2767 [imported] - Vibrio cholerae (strain N16961 se
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: A82037
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, J
hudson, D.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: A82037
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <HEI>
A:Cross-references: GB:AE004342; GB:AE003852; NID:g9657358; PIDN:AAF95906.1; GSPDB:
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2767
A:Map position: 1
C:Superfamily: H+-transporting ATP synthase delta chain

Query Match 13.6%; Score 72.5; DB 2; Length 177;
Best Local Similarity 32.6%; Pred. No. 0.79;
Matches 15; Conservative 11; Mismatches 13; Indels 7; Gaps 1;
Qy 4 VSRDHASLGDSSETLSQTELRKKER-----KKRERKFOANCGID 42
Db 99 LKQHEKKVDVEVISATELSEQRSEIGSKLEQLERKVKVQLNCSD 144

RESULT 2

T32024
hypothetical protein T06D4.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32024
R:Le, T.T.; Goela, D.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid T06D4.
A:Reference number: Z21113
A:Accession: T32024
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-684 <LEF>
A:Cross-references: EMBL:AF016673; PIDN:AAB66123.1; GSPDB:GN00020; CESP:T06D4.4
A:Experimental source: strain Bristol N2; clone T06D4
C:Genetics:
A:Gene: CESP:T06D4.4
A:Map position: 2
A:Introns: 75/3; 121/1; 150/3; 183/1; 227/3; 319/3; 357/3; 408/1; 444/3; 497/1; 62/

Query Match 13.2%; Score 70; DB 2; Length 684;
Best Local Similarity 31.8%; Pred. No. 5.8;
Matches 14; Conservative 9; Mismatches 9; Indels 12; Gaps 1;
QY 23 RKERKKRERKFOA-----NCGIDFIIFWIFWILLF 54
DB 8 KEKERKKAERKFADEKATANKRIINCVPIILIFWILLF 51
RESULT 3
A40470
glucocorticoid-induced receptor precursor, short form RP39 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 20-Apr-2000
C:Accession: A40470
R:Harrigan, M.T.; Campbell, N.F.; Bourgeois, S.
Mol. Endocrinol. 5, 1331-1338, 1991
A:Title: Identification of a gene induced by glucocorticoids in murine T-cells: a potent
A:Reference number: A40470; MUID:92123228; PMID:1663214
A:Accession: A40470
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-381 <HAR>
A:Cross-references: GB:M80610
C:Superfamily: neurokinin 1 receptor

Query Match 12.8%; Score 68; DB 2; Length 381;
Best Local Similarity 22.8%; Pred. No. 5.6;
Matches 29; Conservative 11; Mismatches 41; Indels 46; Gaps 5;
QY 10 SLGDSLTLSQTELRKKRERKFOANCIGDIFIWI---FWILL----- 53
DB 229 TIGDVTTEQYIALR---RKKKTTVKMLVLVVVLFALCWFPPLNCYVLLSSKAIHTNNALY 285
QY 54 FSHHWIQESLLC-----PPSPKEVTCREMLTGGLPWAT 87
DB 286 FAFHWFAMSTCYNPFYICWLNENFRVELKALLSMCQRPKPQEDRLPSPVPSFRVAWTE 345
QY 88 RSHLGR 94
DB 346 KSH-GRR 351

RESULT 4
B40470
glucocorticoid-induced receptor precursor, short form RP23 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 20-Apr-2000
C:Accession: B40470
R:Harrigan, M.T.; Campbell, N.F.; Bourgeois, S.
Mol. Endocrinol. 5, 1331-1338, 1991
A:Title: Identification of a gene induced by glucocorticoids in murine T-cells: a potent
A:Reference number: A40470; MUID:92123228; PMID:1663214
A:Accession: B40470
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-423 <HAR>
A:Cross-references: GB:M80481; GB:M80610; NID:g193516; PID:AAAI7882.1; PID:g460318
C:Superfamily: neurokinin 1 receptor

Query Match 12.8%; Score 68; DB 2; Length 423;
Best Local Similarity 22.8%; Pred. No. 6.2;
Matches 29; Conservative 11; Mismatches 41; Indels 46; Gaps 5;
QY 10 SLGDSLTLSQTELRKKRERKFOANCIGDIFIWI---FWILL----- 53
DB 271 TIGDVTTEQYIALR---RKKKTTVKMLVLVVVLFALCWFPPLNCYVLLSSKAIHTNNALY 327
QY 54 FSHHWIQESLLC-----PPSPKEVTCREMLTGGLPWAT 87
DB 328 FAFHWFAMSTCYNPFYICWLNENFRVELKALLSMCQRPKPQEDRLPSPVPSFRVAWTE 387

QY 88 RSHLGR 94
DB 388 KSH-GRR 393

RESULT 5
D40470
glucocorticoid-induced receptor, long form RP105 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 20-Apr-2000
C:Accession: D40470
R:Harrigan, M.T.; Campbell, N.F.; Bourgeois, S.
Mol. Endocrinol. 5, 1331-1338, 1991
A:Title: Identification of a gene induced by glucocorticoids in murine T-cells: a potent
A:Reference number: A40470; MUID:92123228; PMID:1663214
A:Accession: D40470
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-443 <HAR>
A:Cross-references: GB:M80160
C:Superfamily: neurokinin 1 receptor

Query Match 12.8%; Score 68; DB 2; Length 443;
Best Local Similarity 22.8%; Pred. No. 6.4;
Matches 29; Conservative 11; Mismatches 41; Indels 46; Gaps 5;
QY 10 SLGDSLTLSQTELRKKRERKFOANCIGDIFIWI---FWILL----- 53
DB 291 TIGDVTTEQYIALR---RKKKTTVKMLVLVVVLFALCWFPPLNCYVLLSSKAIHTNNALY 347
QY 54 FSHHWIQESLLC-----PPSPKEVTCREMLTGGLPWAT 87
DB 348 FAFHWFAMSTCYNPFYICWLNENFRVELKALLSMCQRPKPQEDRLPSPVPSFRVAWTE 407
QY 88 RSHLGR 94
DB 408 KSH-GRR 413

RESULT 6
C40470
glucocorticoid-induced receptor, long form RP82 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 20-Apr-2000
C:Accession: C40470
R:Harrigan, M.T.; Campbell, N.F.; Bourgeois, S.
Mol. Endocrinol. 5, 1331-1338, 1991
A:Title: Identification of a gene induced by glucocorticoids in murine T-cells: a potent
A:Reference number: A40470; MUID:92123228; PMID:1663214
A:Accession: C40470
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-491 <HAR>
A:Cross-references: GB:M80610
C:Superfamily: neurokinin 1 receptor

Query Match 12.8%; Score 68; DB 2; Length 491;
Best Local Similarity 22.8%; Pred. No. 7.1;
Matches 29; Conservative 11; Mismatches 41; Indels 46; Gaps 5;
QY 10 SLGDSLTLSQTELRKKRERKFOANCIGDIFIWI---FWILL----- 53
DB 339 TIGDVTTEQYIALR---RKKKTTVKMLVLVVVLFALCWFPPLNCYVLLSSKAIHTNNALY 395
QY 54 FSHHWIQESLLC-----PPSPKEVTCREMLTGGLPWAT 87
DB 396 FAFHWFAMSTCYNPFYICWLNENFRVELKALLSMCQRPKPQEDRLPSPVPSFRVAWTE 455
QY 88 RSHLGR 94
DB 456 KSH-GRR 461

```

RESULT 9
B71600
variant-specific surface protein 1 homolog PF01055c - malaria parasite (Plasmodium
N;Alternate names: erythrocyte membrane binding protein 1 (EMP1)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71600
R; Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin,
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith,
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: B71600
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2197 <GAR>
A;Cross-references: GB:AE001434; GB:AE001362; NID:g3845341; PIDN:AAC71996.1; PID:g383
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PF01055c

Query Match 12.5%; Score 66.5; DB 2; Length 2197;
Best Local Similarity 24.3%; Pred. No. 47;
Matches 21; Conservative 15; Mismatches 50; Indels 19; Gaps 4;

QY 1 EVEYSRDHASLGDSFTLSQTELRLK-----KRRKKRRKFOAN-----CGIDFI 44
Db 602 EVQVKDKDN-GEDEDEEDVDKVKAGGLCILENKKHESRNNSSNEPEQFTFHDF 660
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 45 IFWTFWILLFHHHTQESLLCPSPKEVTCREMLTGCLPWARTRSHLGRKK 95
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 661 YFWIGRFLNDSMYWRGVNSCINPNRKKRCRNECKDDC--GCFKEWIGKKK 709

RESULT 10
T49635
hypothetical protein B5022.170 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C;Accession: T49635
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyaka
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49635
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-132 <SCH>
A;Cross-references: EMBL:AL355932; GSPDB:GN00116; NCSP:B5022.170
A;Experimental source: BAC clone B5022; strain OR74A
C;Genetics:
A;Gene: NCSP:B5022.170
A;Map position: 6
C;Superfamily: Neurospora crassa hypothetical protein B5022.170

Query Match 12.4%; Score 66; DB 2; Length 132;
Best Local Similarity 25.0%; Pred. No. 3.3;
Matches 21; Conservative 14; Mismatches 31; Indels 18; Gaps 3;

QY 9 ASLGSDETLSQTELRLKRRKKRKEFOANGCID-----FIIFIW---FWILLFHHW 58
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 SSVTTTIRWGRKEKKKKRKRKEQNAKDGTTVADGORSVVYGVWLGFWPSLHHAA 96

QY 59 IQE-----SLLCPPSPKEVTC 74
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 97 VHELSTTTVHHQLNPPAPLSPMPC 120

RESULT 11
F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201

```

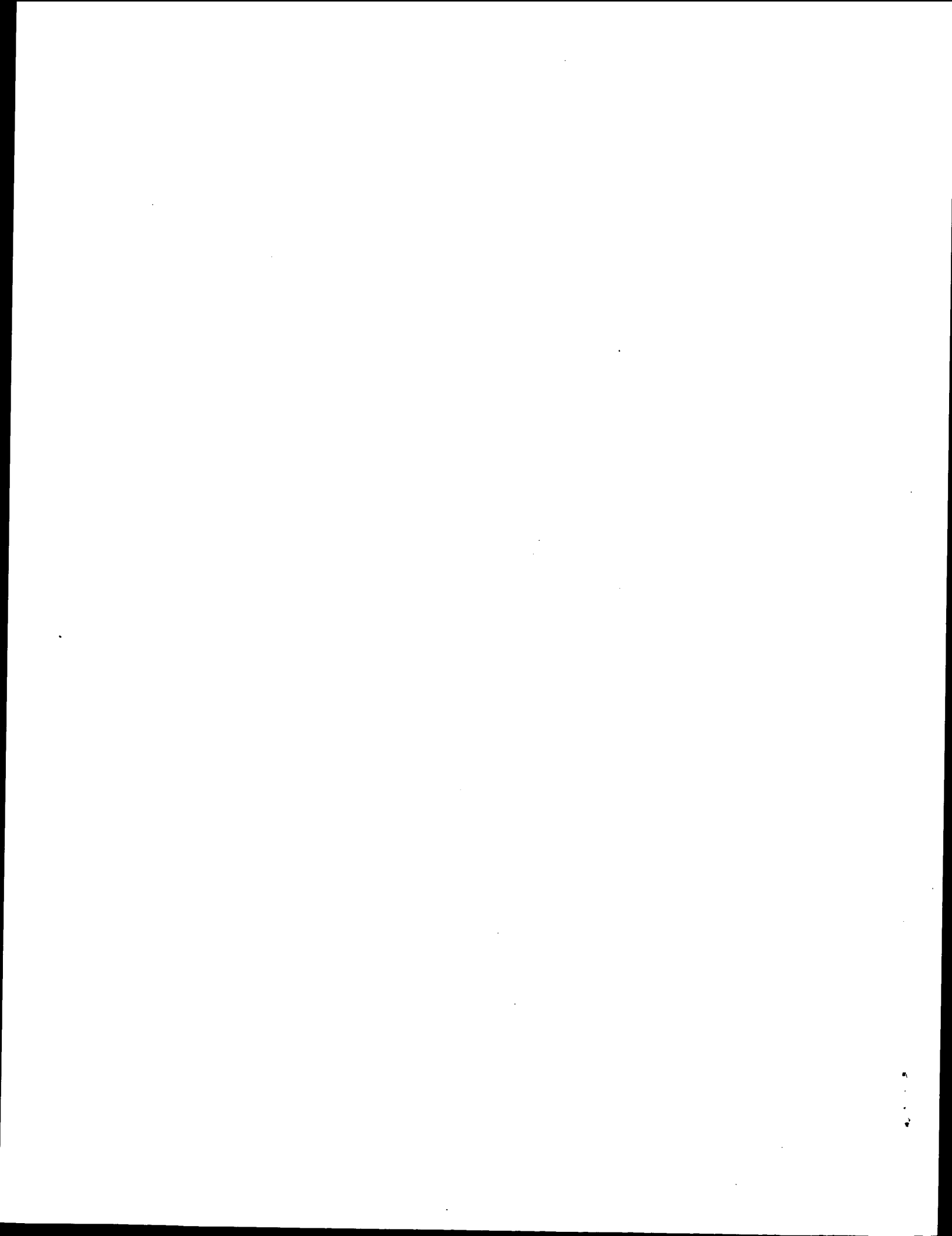
[illegible]

Wed Jul 9 15:44:29 2003

C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology
F:312-480/Domain: cytochrome P450 homology <p45>

Query Match 12.2%; Score 65; DB 2; Length 514;
Best Local Similarity 20.0%; Pred. No. 17;
Matches 34; Conservative 21; Mismatches 33; Indels 82; Gaps 9;
QY 6 RDHASLGDSSETLSQ--TELK---KKERKKRE-----RKFO 36
Db 246 KKYFNIGSEKRLKEAVTEVRSFAKKLVREKKRELEKSSLETEDMLSRFLSSGHSDDEV 305
QY 37 ANCGIDFI-----IFWIFWIL-----LFSHH 57
Db 306 ADIVISFILAGKDTTSAALTWFFWLLWKNPRVEEIVNELSKKSELVMVYDEVKEMVYTHA 365
QY 58 WIOESL-LGPPSPKEVTCREMLTGGCLP--WATRS-----HLGRRK 95
Db 366 ALSESMRLYPPVPMD--SKEAVNDVLPDGVVVKGTIVTYHYVYANGRMK 413

Search completed: July 9, 2003, 15:09:44
Job time : 46.7788 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 12.0177 Seconds
(without alignments)
334.773 Million cell updates/sec

Title: US-09-854-133-586
Perfect score: 532
Sequence: 1 EVEVSRDHSALGDSFTLSQT.....LTGGCLPWATRSHLGRKKS 97

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	68	12.8	1 GP72_MOUSE	P30731 mus musculus
2	67	12.6	1 SH2C_HUMAN	P28335 homo sapien
3	65.5	12.3	1 YNFI_ECOLI	P76174 escherichia
4	65	12.2	1 C941_VICSA	O81117 vicia sativ
5	63	11.8	1 APJ_XENLA	P79960 xenopus lae
6	62.5	11.7	1 ATPD_VIBAL	P12987 vibrio algi
7	62.5	11.7	1 BPHY_PSEAE	O9hwr3 pseudomonas
8	62	11.7	1 TRHY_SHEEP	P22793 ovis aries
9	61.5	11.6	1 ACDL_CLOPE	O8xmy2 clostridium
10	61.5	11.6	1 SH7_HUMAN	P34969 homo sapien
11	61.5	11.6	1 RSI_RHIME	P14129 rhizobium m
12	61.5	11.6	1 POLG_HRV89	P07210 human rhino
13	61	11.5	1 PPAL_PEDAC	P29430 pedicoccus
14	60.5	11.4	1 ETS2_MOUSE	P15037 mus musculus
15	60	11.3	1 VEA_HPV11	P04016 human papil
16	60	11.3	1 VEA_HPV6C	P20969 human papil
17	60	11.3	1 SH2C_MOUSE	P34968 mus musculus
18	60	11.3	1 CBP5_YEAST	P33322 saccharomyc
19	59.5	11.2	1 MYOG_CHICK	P17920 gallus gall
20	59.5	11.2	1 ANR2_HUMAN	O9gzw1 homo sapien
21	59.5	11.2	1 SH7_CAVPO	P47148 saccharomyc
22	59.5	11.2	1 SH7_MOUSE	P32304 mus musculus
23	59.5	11.2	1 SH7_RAT	P32305 rattus norv
24	59.5	11.2	1 HJH1_HUMAN	P35367 homo sapien
25	59.5	11.2	1 ACM3_CHICK	P49578 gallus gall
26	59.5	11.2	1 CBP_MOUSE	P45481 mus musculus
27	59.5	11.2	1 CBP_HUMAN	O92793 homo sapien
28	59	11.1	1 YJ81_YEAST	P47148 saccharomyc
29	59	11.1	1 SH2C_RAT	O89099 rattus norv
30	59	11.1	1 NBL4_BRARE	O57457 brachydanio
31	59	11.1	1 ICAL_RAT	P27321 rattus norv
32	58.5	11.0	1 SRYB_DROME	P07665 drosophila
33	58.5	11.0	1 TRFE_HUMAN	P02787 homo sapien

ALIGNMENTS

RESULT 1									
GP72_MOUSE									
ID	GP72_MOUSE	STANDARD;	PRT;	423 AA.					
AC	P30731;								
DT	01-APR-1993 (Rel. 25, Created)								
DT	01-NOV-1995 (Rel. 32, Last sequence update)								
DT	16-OCT-2001 (Rel. 40, Last annotation update)								
DE	Probable G protein-coupled receptor GPR72 precursor (Glucocorticoid-								
DE	induced receptor).								
GN	GPR72 OR GIR.								
OS	Mus musculus (Mouse).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
OX	NCBI_TaxID=10090;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RA	MEDLINE=92123228; PubMed=1663214;								
RA	Harrigan M.F., Campbell N.F., Bourgeois S.;								
RT	"Identification of a gene induced by glucocorticoids in murine								
RT	T-cells: a potential G protein-coupled receptor.";								
RL	Mol. Endocrinol. 5:1331-1338(1991).								
CC	-!- FUNCTION: ORPHAN RECEPTOR. COULD BE A NEUROPEPTIDE Y RECEPTOR.								
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein.								
CC	-!- ALTERNATIVE PRODUCTS: MULTIPLE FORMS OF GIR MAY ARISE BY								
CC	ALTERNATIVE SPLICING.								
CC	-!- TISSUE SPECIFICITY: BRAIN, THYMUS AND SPLEEN.								
CC	-!- INDUCTION: BY GLUCOCORTICIDS AND CAMP; IN T-CELLS.								
CC	-!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.								
CC	-----								
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration								
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -								
CC	the European Bioinformatics Institute. There are no restrictions on its								
CC	use by non-profit institutions as long as its content is in no way								
CC	modified and this statement is not removed. Usage by and for commercial								
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ ;								
CC	or send an email to license@isb-sib.ch).								
CC	-----								
CC	EMBL; M80481; AAA17882.1; -								
DR	MGD; MGI:95712; Gir.								
DR	InterPro; IPR000276; GPCR_Rhodpsn.								
DR	Pfam; PF00001; 7tm1.1;								
DR	PRINTS; PR00237; GPCRHHODPSN								
DR	PROSITE; PS00237; G_PROTEIN_REC_F1_1; 1.								
DR	PROSITE; PS00262; G_PROTEIN_REC_F1_2; 1.								
KW	G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;								
FT	Alternative splicing.								
FT	SIGNAL 1 17								
FT	CHAIN 18 423								
FT	POTENTIAL.								
FT	PROBABLE G PROTEIN-COUPLED RECEPTOR								
FT	GPR72.								
FT	DOMAIN 18 71								
FT	EXTRACELLULAR (POTENTIAL).								
FT	TRANSMEM 72 92								
FT	1 (POTENTIAL).								
FT	DOMAIN 93 107								
FT	CYTOPLASMIC (POTENTIAL).								
FT	TRANSMEM 108 129								
FT	2 (POTENTIAL).								
FT	DOMAIN 130 145								
FT	EXTRACELLULAR (POTENTIAL).								
FT	DOMAIN 146 167								
FT	3 (POTENTIAL).								
FT	DOMAIN 168 186								
FT	CYTOPLASMIC (POTENTIAL).								

P56566 mus musculus
P53724 saccharomyc
O14966 homo sapien
Q28509 macaca mula
Q40986 saccharomyc
P11267 simian immu
P91620 drosophila
Q9wv06 mus musculus
P38487 bacillus sp
Q8vc56 mus musculus
P38361 saccharomyc

FT TRANSMEM 187 208 4 (POTENTIAL).
 FT DOMAIN 238 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 239 5 (POTENTIAL).
 FT DOMAIN 260 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 261 6 (POTENTIAL).
 FT DOMAIN 294 315 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 316 327 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 328 348 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 349 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 144 224 BY SIMILARITY.
 FT VARSPLIC 130 171 MISSING (IN ISOFORM RP39).
 FT VARSPLIC 171 171 Q -> QGLTAIAVDHGLELQKVRPRGSGELRSPSVTF
 FT VARSPLIC 171 171 VPSLCPALFTCKRPWDFQESQSLHDTLFPPLP (IN
 FT VARSPLIC 171 171 ISOFORM RP82).
 FT VARSPLIC 171 171 Q -> QRPWDFQESQSLHDTLFPPLP (IN ISOFORM
 FT VARSPLIC 171 171 RP105).
 FT SEQUENCE 423 AA; 48136 MW; 3ACE43452BF15391 CRC64;
 SO SEQUENCE 423 AA; 48136 MW; 3ACE43452BF15391 CRC64;
 Query Match 12.8%; Score 68; DB 1; Length 423;
 Best Local Similarity 22.8%; Pred. No. 3.8;
 Matches 29; Conservative 11; Mismatches 41; Indels 46; Gaps 5;
 QY 10 SLGDSLTLSQTELRKKRKKRERKFOANGDIFIFWI---FWILL----- 53
 DB 271 TIGDVTTEQYALR---RKKKTVMKLVLVVLFALCWFPLNCYVLLLSKAIHTNNALY 327
 QY 54 FSHWIOQLSLC-----PPSPKEVTCREMLTGGLPWAT 87
 DB 328 FAFHWANSSCYNPFYICWLNENFVELKALLSMCQRPKQPDRLPSPVSRVAWTE 387
 QY 88 RSHLGR 94
 DB 388 KSH-GRR 393

RESULT 2
 SH2C_HUMAN STANDARD; PRT; 458 AA.
 ID AC 5H2C_HUMAN
 AD P28335; Q9NP28;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 5-hydroxytryptamine 2C receptor (5-HT-2C) (Serotonin receptor)
 DE (SHT-1C).
 GN HTR2C OR HTRIC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=92105767; PubMed=1722404;
 RA Saltzman A.G., Morse B., Whitman M.M., Ivanschchenko Y., Jaye M.,
 RA Felder S.;
 RT "Cloning of the human serotonin 5-HT2 and 5-HT1C receptor subtypes.";
 RL Biochem. Biophys. Res. Commun. 181:1469-1478(1991).
 RN [2]
 RP SEQUENCE FROM N.A., AND FUNCTION.
 RC TISSUE=Placenta, and Hippocampus;
 RX MEDLINE=95203331; PubMed=7895773;
 RA Stam N.J., Vanderheyden P., Van Alebeek C., Klomp J., De Boer T.,
 RA Van Delft A.M.L., Olijve W.;
 RT "Genomic organisation and functional expression of the gene encoding
 the human serotonin 5-HT2C receptor.";
 RL Eur. J. Pharmacol. 269:339-348(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97001158; PubMed=8812491;
 RA Xie E., Zhao L., Levine A.J., Shenk T., Chang L.-S.;

RT "The human serotonin 5-HT2C receptor: complete cDNA, genomic
 structure, and alternatively spliced variant.";
 RL Genomics 35:551-561(1996).
 RN [4]
 RP SEQUENCE FROM N.A., AND RNA EDITING.
 RC TISSUE=Brain;
 RX MEDLINE=99127198; PubMed=9928237;
 RA Niswender C.M., Sanders-Bush E., Emeson R.B.;
 RT "Identification and characterization of RNA editing events within the
 5-HT2C receptor.";
 RL Ann. N.Y. Acad. Sci. 861:38-48(1998).
 RN [5]
 RP SEQUENCE OF 1-116 FROM N.A.
 RA Kalicki J., Mead K.;
 RL Submitted (Nov-1998) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP VARIANT SER-23
 RX MEDLINE=96044432; PubMed=7557992;
 RA Lappalainen J., Zhang L., Dean M., Oz M., Ozaki N., Yu D.,
 RA Virkkunen M., Weight F., Linnoila M., Goldman D.;
 RT "Identification, expression, and pharmacology of a Cys23-Ser23
 substitution in the human 5-HT2c receptor gene (HTR2C).";
 RL Genomics 27:274-279(1995).
 CC -!- FUNCTION: THIS IS ONE OF THE SEVERAL DIFFERENT RECEPTORS FOR
 5-HYDROXYTRYPTAMINE (SEROTONIN), A BIOGENIC HORMONE THAT FUNCTIONS
 AS A NEUROTRANSMITTER, A HORMONE, AND A MITOGEN. THIS RECEPTOR
 MEDIATES ITS ACTION BY ASSOCIATION WITH G PROTEINS THAT ACTIVATE
 A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- POLYMORPHISM: POSITION 23 IS POLYMORPHIC; THE FREQUENCIES IN
 UNRELATED CAUCASIANS ARE 0.87 FOR CYS AND 0.13 FOR SER.
 CC -!- MISCELLANEOUS: RNA EDITING GENERATES RECEPTOR ISOFORMS THAT DIFFER
 IN THEIR ABILITY TO INTERACT WITH THE PHOSPHOLIPASE C SIGNALING
 CASCADE IN A TRANSFECTED CELL LINE, SUGGESTING THAT THIS RNA
 PROCESSING EVENT MAY CONTRIBUTE TO THE MODULATION OF SEROTONERGIC
 NEUROTRANSMISSION IN THE CENTRAL NERVOUS SYSTEM.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC STRONGEST TO THE OTHER 5HT-2 SUBTYPE RECEPTORS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M81778; AAA60317.1; -;
 CC EMBL; X80763; CAB59978.1; -;
 CC EMBL; U49516; AAB40898.1; -;
 CC EMBL; AF208053; AAF35842.1; -;
 CC EMBL; AC004822; AAC71658.1; -;
 CC PIR; JS0616; JS0616.
 CC Genew; HGNC:5295; HTR2C.
 CC MIM; 312861; -;
 CC InterPro; IPR000276; GPCR_Rhodpsn.
 CC Pfam; PF00001; 7tm_1; 1.
 CC PRINTS; PR00237; GPCR_RHODPSN.
 CC PROSITE; PS00237; G_PROTEIN_RECP_F1_1; 1.
 CC PROSITE; PS00262; G_PROTEIN_RECP_F1_2; 1.
 CC G-protein coupled receptor; Transmembrane; Glycoprotein; Polymorphism;
 Multigene family; RNA editing.
 FT DOMAIN 1 52 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 53 78 1 (POTENTIAL).
 FT DOMAIN 79 89 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 90 110 2 (POTENTIAL).
 FT DOMAIN 111 127 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 128 150 3 (POTENTIAL).
 FT DOMAIN 151 170 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 171 193 4 (POTENTIAL).
 FT DOMAIN 194 213 5 (POTENTIAL).
 FT TRANSMEM 214 235 5 (POTENTIAL).
 FT DOMAIN 236 311 CYTOPLASMIC (POTENTIAL).

[illegible]

RESULT 3

YNFI_ECOLI	YNFI_ECOLI	STANDARD;	PRT; 207 AA.
ID	YNFI_ECOLI		
AC	P6174; P7270;		
DT	15-JUL-1998 (Rel. 36, Created)		
DC	15-JUL-1998 (Rel. 36, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Hypothetical protein ynfI precursor.		
DE	YNFI OR B1591 OR Z2581 OR ECS2297.		
GN	Escherichia coli, and		
OS	Escherichia coli, and		
OS	Escherichia coli O157:H7.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Escherichia.		
OC	NCBI_TaxID=562, 83334;		
EN	[1]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN-K12 / MG1655;		
RP	MEDLINE=97426617; PubMed=9278503;		
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,		
RA	Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,		
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,		
RA	Mau B., Shao Y.;		
RA	"The complete genome sequence of Escherichia coli K-12.";		
RL	Science 277:1453-1474(1997).		
[2]			
RP	SEQUENCE FROM N.A.		
RP	STRAIN-K12;		
RP	MEDLINE=97251357; PubMed=9097039;		
RA	Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,		
RA	Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,		
RA	Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,		
RA	Motomura K., Nakade S., Nakamura Y., Nashimoto H., Nishio Y.,		
RA	Osahima T., Saito N., Sanpei G., Seki Y., Sivasundaram S.,		
RA	Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,		
RA	Yamamoto Y., Horiuchi T.;		
RT	"A 570-kb DNA sequence of the Escherichia coli K-12 genome		
RT	corresponding to the 28.0-40.1 min region on the linkage map.";		
RL	DNA Res. 3:363-377(1996).		
[3]			
RP	SEQUENCE FROM N.A.		
RP	STRAIN=O157:H7 / EDL933 / ATCC 700927;		
RC	MEDLINE=21074935; PubMed=11206551;		
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,		
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,		

Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
Apodaca J., Anantharaman I.S., Lin J., Yen G., Schwartz D.C.,
Welch R.A., Blattner F.R.;
"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
Nature 409:529-533(2001).
[4]
SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RMD 0509952;
RC MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Tida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
Kuhara S., Shiba T., Hattori M., Shinagawa H.;
"Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12";
DNA Res. 8:11-22(2001).
RT
CC -1- SIMILARITY: TO H.INFLUENZAЕ H1044.
CC -1- SIMILARITY: BELONGS TO THE YCDY/YNFI FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AE000254; AAC74663.1; -;
DR EMBL; D90801; BAA15315.1; ALT_INIT.
DR EMBL; D90802; BAA1325.1; ALT_INIT.
DR EMBL; AE005382; AAG56578.1; -;
DR EMBL; AP002557; BAB35720.1; -;
DR EcoGene; EGI3847; ynfI.
KW Hypothetical protein; Signal; Complete proteome.
KW
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 207 HYPOTHETICAL PROTEIN YNFI.
FT SEQUENCE 207 AA; 23646 MW; 458DDBF76D16EA89 CRC64;
SQ
Query Match 12.3%; Score 65.5; DB 1; Length 207;
Best Local Similarity 28.7%; Pred. No. 3.4;
Matches 25; Conservative 14; Mismatches 33; Indels 15; Gaps 4
QY 2 VEVSERDASIGDSETLSQTELRKKRKKRKKRKKFOANGIDFIEFWIFLFSHHWIOE 61
DB 93 VMLDRESVLFGDS-TLA---LRQWREKGIQFMQKQNEPDH-----FGSLLMAANLAE 143
QY 62 SLLCPPSPKEVTCREMLTGGGCLPWATR 88
DB 144 -----NGRQTECEELLAWHLFPWSTR 164
RESULT 4
C941_VICSA
ID C941_VICSA STANDARD; PRT; 514 AA.
AC O81117;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome P450 94A1 (EC 1.14.-.-) (P450-dependent fatty acid
DE omega-hydroxylase).
DE CYP94A1 OR VAGH111.
GN Vicia sativa (Spring vetch) (Tare).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Vicia.
OX NCBI_TaxID=3908;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Seedling.
RX MEDLINE=98264856; PubMed=9601090;
RA Tijet N., Helvig C., Pinot F., Le Bouquin R., Lesot A., Durst F.,
RA Salauen J.-P., Benveniste I.;

FT	REPEAT	449	476	1-2.
FT	REPEAT	477	504	1-3.
FT	REPEAT	505	532	1-4.
FT	REPEAT	533	560	1-5.
FT	REPEAT	561	588	1-6.
FT	REPEAT	589	616	1-7.
FT	REPEAT	617	644	1-8.
FT	REPEAT	645	678	1-9.
FT	REPEAT	679	706	1-10.
FT	REPEAT	707	742	1-11.
FT	REPEAT	743	771	1-12.
FT	REPEAT	772	796	1-13.
FT	REPEAT	797	832	1-14.
FT	DOMAIN	938	1507	23 X 23 AA APPROXIMATE TANDEM REPEATS.
FT	REPEAT	938	961	2-1.
FT	REPEAT	962	985	2-2.
FT	REPEAT	986	1021	2-3.
FT	REPEAT	1022	1044	2-4.
FT	REPEAT	1045	1067	2-5.
FT	REPEAT	1068	1090	2-6.
FT	REPEAT	1091	1121	2-7.
FT	REPEAT	1122	1144	2-8.
FT	REPEAT	1145	1167	2-9.
FT	REPEAT	1168	1197	2-10.
FT	REPEAT	1198	1227	2-11.
FT	REPEAT	1228	1250	2-12.
FT	REPEAT	1251	1273	2-13.
FT	REPEAT	1274	1296	2-14.
FT	REPEAT	1297	1319	2-15.
FT	REPEAT	1320	1342	2-16.
FT	REPEAT	1343	1368	2-17.
FT	REPEAT	1369	1391	2-18.
FT	REPEAT	1392	1416	2-19.
FT	REPEAT	1417	1439	2-20.
FT	REPEAT	1440	1461	2-21.
FT	REPEAT	1462	1484	2-22.
FT	REPEAT	1485	1507	2-23.
FT	VARIANT	1145	1197	MISSING (IN SHORT FORM).
FT	VARIANT	1251	1273	MISSING (IN SHORT FORM).
FT	CONFLICT	1399	1399	E -> G (IN REF. 2).
SQ	SEQUENCE	1549 AA; 201173 MW; E72FB9FF1326E34E CRC64;		

Query Match 11.7%; Score 62; DB 1; Length 1549;

Best Local Similarity 37.1%; Pred. No. 69;

Matches 13; Conservative 9; Mismatches 13; Indels 0; Gaps 0;

Oy	1	EVEYSRHOASLGDSSETLSQTLRKKKRKKRKF	35
		: :	
Db	418	EVELQREERLQREEEQLQREERKRQREKQY	452

RESULT 9

ACD1_CLOPE

ID	ACD1_CLOPE	STANDARD;	PRT;	218 AA.
AC	Q8XMY2;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Putative acyl carrier protein phosphodiesterase 1 (EC 3.1.4.14) (ACP phosphodiesterase 1).			
DE	CPE0556			
GN	Clostridium perfringens.			
OS	Clostridium perfringens.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
ON	NCBI_TaxID=1502;			
OX	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=13 / Type A;			
RX	PubMed=11792842;			
RA	Shimizu T., Ohkawa K., Hirakawa H., Ohshima K., Yamashita A.,			
RA	Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,			
RT	"Complete genome sequence of Clostridium perfringens, an anaerobic			
RT	flesh-eater."			


```
Db 301 EVCNLSRLKHERKNISIPKREKQKATTLGIIVGATVCWLPFF-----LLS----- 347
QY 71 EVCNLSRLKHERKNISIPKREKQKATTLGIIVGATVCWLPFF-----LLS----- 347
Db 348 --TARPFICGTCSCIPLWVETFL 370

RESULT 11
RSL_RHIME
ID RSL_RHIME STANDARD; PRT; 568 AA.
AC P14129;
DT 01-JAN-1990 (Rel. 13, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S1.
GN RPSA OR R00255 OR SMC00335.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / SU47;
RX MEDLINE=88217521; PubMed=3368316;
RA Schmier J., Thamm S., Lurz R., Hussain A., Faist G.,
RA Dobrinski B.
RT "Cloning and characterization of a gene from Rhizobium meliloti 2011
RT coding for ribosomal protein S1."
RL Nucleic Acids Res. 16:3075-3089(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galberget F.,
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -!- FUNCTION: BINDS MRSA; THUS FACILITATING RECOGNITION OF THE
CC INITIATION POINT. IT IS NEEDED TO TRANSLATE MRNA WITH A SHORT
CC SHINE-DALGARNO (SD) PURINE-RICH SEQUENCE (BY SIMILARITY)
CC -!- SIMILARITY: BELONGS TO THE S1P FAMILY OF RIBOSOMAL PROTEINS.
CC -!- SIMILARITY: CONTAINS 6 S1 MOTIF DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
```

```
FT DOMAIN 459 530 S1 MOTIF 6.
FT CONFLICT 88 88 A -> R (IN REF. 1).
FT CONFLICT 153 162 LMHNQPFEL -> ADAQPAALRN (IN REF. 1).
FT CONFLICT 235 235 N -> K (IN REF. 1).
FT CONFLICT 241 241 L -> Q (IN REF. 1).
FT CONFLICT 551 551 A -> R (IN REF. 1).
SQ SEQUENCE 568 AA; 62640 MW; AB0858204273A7B8 CRC64;

Query Match
Best Local Similarity 11.6%; Score 61.5; DB 1; Length 568;
Matches 17; Conservative 12; Mismatches 16; Indels 1; Gaps 1;

QY 1 EVEYSRDHASLGDSETLSQTELKRRKKRKKRKFQANGCIDFIIF 46
||| | :||: ||: |::| :|::| :|::| :|::| :|::|
Db 75 EVYVERIENALGEA-VLSREKARREESQRLVKEFGEGRVEGIIF 119

RESULT 12
POLG_HRV89
ID POLG_HRV89 STANDARD; PRT; 2164 AA.
AC P07210; Q82096; Q82097; Q82098; Q82099; Q82100; Q82101; Q82102;
AC Q82103; Q82104; Q82105;
DT 01-APR-1988 (Rel. 07, Created)
DT 15-APR-2002 (Rel. 41, Last sequence update)
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C, P3A; Genome-linked protein VP6; Picornain 3C
DE (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D
DE (EC 2.7.7.48)].
OS Human rhinovirus 89 (HRV-89).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Rhinovirus.
OX NCBI_TaxID=12132;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87204179; PubMed=3033653;
RA Duechler M., Skern T., Sommergruber W., Neubauer C., Gruendler P.,
RA Fogy I., Blaas D., Kuechler E.;
RT "Evolutionary relationships within the human rhinovirus genus:
RT comparison of serotypes 89, 2, and 14."
RL Proc. Natl. Acad. Sci. U.S.A. 84:2605-2609(1987).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-I-Gly bond in the
CC poliovirus polyprotein. In other picornavirus reactions Glu may be
CC substituted for Gln, and Ser or Thr for Gly.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
```


[illegible]


```

AC Q9BJ7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Glucocorticoid induced receptor (Fragment).
GN GPR83 OR GPR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SV;
RX MEDLINE=20515601; PubMed=11060465;
RA De Moerloose L., Williamson J., Liners F., Perret J., Parmentier M.;
RT "Cloning and chromosomal mapping of the mouse and human genes encoding
RL the orphan glucocorticoid-induced receptor (GPR83).";
DR EMBL; Y19225; CAC19040.1; -.
DR EMBL; Y19225; CAC19040.1; JOINED.
DR EMBL; Y19227; CAC19040.1; JOINED.
DR MGD; MGI:95712; Gpr83.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCR_RHODOPS.
DR PROSITE; PS00237; G_PROTEIN_RECEP_FL_1; UNKNOWN_1.
DR PROSITE; PS0262; G_PROTEIN_RECEP_FL_2; 1.
KW Receptor.
FT NON_TER 463 463
SQ SEQUENCE 463 AA; 52804 MW; BBC7E9F8B717CE0 CRC64;

Query Match 12.8%; Score 68; DB 11; Length 463;
Best Local Similarity 22.8%; Pred. No. 7.6;
Matches 29; Conservative 11; Mismatches 41; Indels 46; Gaps 5;

QY 10 SLGDSFSLTSLQTELKRRKKRRKFKQANGCIDFIIFWI----FWILL----- 53
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 328 TIGDVTTEQYLALR---RKKKTKVLMVLVVLVLCWFLPNCVLLSSKAIHTNNALY 384
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 54 FSHHWQESLLC-----PPSPKVEVTCREMLTGGCLPWAT 87
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 385 FAFHWFAMSTCYNPFIYCNLENFRVELKALLSMCQRPKPQEDRLSPVPSFRVAVTE 444
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 88 RSHLGR 94
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 445 KSH-GRR 450

Query Match 12.7%; Score 67.5; DB 5; Length 411;
Best Local Similarity 25.3%; Pred. No. 7.8;
Matches 23; Conservative 12; Mismatches 33; Indels 23; Gaps 5;

QY 7 DHASLGDSETLTSLQTELKRRKKRRKFKQANGC-----IDFI-IFWTFWILLFHHWITQ 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 278 DKGSSNSNRVSRQKRRKGRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRR 328
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 ESLLCPSPKVEVTCREMLTGGCLPWATRSHL 91
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 329 ---VAPPGYQAYYCH---GEC-PFPLADHL 351
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
Q18657
ID Q18657 PRELIMINARY; PRT; 1851 AA.
AC Q18657;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE C46C2.1 protein.
DE C46C2.1.
GN C46C2.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.

```

```

Query Match 12.7%; Score 67.5; DB 15; Length 169;
Best Local Similarity 20.4%; Pred. No. 3.3;
Matches 23; Conservative 19; Mismatches 34; Indels 37; Gaps 4;

QY 3 EVSRDHASLGDSE---TSLQTELKRRKKRRKFKQ-----NCGIDPIIF 46
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 48 DIRKAHCNLTSTKWANTLSQIAKRLKEQSGNKTVFNRRSSGGDPEIVMHSFNCGGEP--F 105
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 47 WIFWILLFHHW-----IQESLLCPSPKVEVTCREMLTG 80
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 106 YCNTTQFNSTWNGTVTLPCRKQIVNMQEVGKAMTAPPYIQGITCSSNITG 158
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 14
Q9U418
ID Q9U418 PRELIMINARY; PRT; 411 AA.
AC Q9U418;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Bone morphogenetic protein 2/4.
GN BMP2/4.
OS Branchiostoma belcheri (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7741;
RN [1]
RP SEQUENCE FROM N.A.
RA Yasui K., Saiga H., Uemura M., Semba I.;
RT "Early body formation and expression pattern of genes encoding
RT secreted proteins, BbBMP2/4, BbWnt7, and BbWnt8 in lancelets.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
DR EMBL; AF206325; AAF19841.1; -.
DR HSSP; P12643; 3BMP.
DR InterPro; IPR002405; Inhibin_alpha.
DR InterPro; IPR001839; TGFb.
DR InterPro; IPR001111; TGFb_N.
DR Pfam; PF000019; TGF-beta; 1.
DR Pfam; PF00688; TGFb_propeptide; 1.
DR PRINTS; PR00669; INHIBINA.
DR PRODOM; PD000357; TGFb; 1.
DR SMART; SM00204; TGFb; 1.
DR PROSITE; PS00250; TGF_BETA_1; 1.
KW Glycoprotein.
SQ SEQUENCE 411 AA; 46929 MW; F579898060F18355 CRC64;

```

```

Query Match 12.7%; Score 67.5; DB 5; Length 411;
Best Local Similarity 25.3%; Pred. No. 7.8;
Matches 23; Conservative 12; Mismatches 33; Indels 23; Gaps 5;

QY 7 DHASLGDSETLTSLQTELKRRKKRRKFKQANGC-----IDFI-IFWTFWILLFHHWITQ 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 278 DKGSSNSNRVSRQKRRKGRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRR 328
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 ESLLCPSPKVEVTCREMLTGGCLPWATRSHL 91
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 329 ---VAPPGYQAYYCH---GEC-PFPLADHL 351
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
Q18657
ID Q18657 PRELIMINARY; PRT; 1851 AA.
AC Q18657;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE C46C2.1 protein.
DE C46C2.1.
GN C46C2.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.

```

```

OX  NCBI_TaxID=6239;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  McMurray A.A.;
RL  Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=99069613; PubMed=9851916;
RA  none;
RT  "Genome sequence of the nematode C.elegans; A platform for
RT  investigating biology.";
RL  Science 282:2012-2018(1998).
CC  -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR  EMBL; Z68296; CAA92591.1; -
DR  InterPro; IPR000719; Euk_pkinase.
DR  InterPro; IPR002290; Ser_thr_pkinase.
DR  Pfam; PF00069; pkinase; 1.
DR  PRINTS; PR01574; TUBBYPROTEIN.
DR  ProDom; PD000001; Euk_pkinase; 1.
DR  PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
DR  PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW  ATP-binding; Serine/threonine-protein kinase; Transferase.
SQ  SEQUENCE 1851 AA; 199918 MW; 71BD78818EF6A5A5 CRC64;

Query Match      12.6%; Score 67; DB 5; Length 1851;
Best Local Similarity 33.3%; Pred. No. 38;
Matches 14; Conservative 9; Mismatches 19; Indels 0; Gaps 0;

Qy  2  VEVSRDHASLGDSETLSQTELKRRKKRRKFKQANGCIDF 43
    ||: ||| | | | | | | | | | | | | | | | |
Db  617 VEIKNRDADLNDLNVIEIQMLRVYDEKKRKYRFFKENEGLQF 658

```

Search completed: July 9, 2003, 15:06:30
Job time : 46.7788 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 8.92035 Seconds
(without alignments)
239.005 Million cell updates/sec

Title: US-09-854-133-587
Perfect score: 98
Sequence: 1 FOANCIGDIFIFWIFW 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	98	100.0	16	22	AAE13851 Human T cell epito
2	50	51.0	504	21	AA54085 Enzyme EPS1 invol
3	50	51.0	504	21	AA54378 Amino acid sequen
4	47	48.0	207	20	AAW99067 Streptococcus pneu
5	47	48.0	326	20	AAW99066 Streptococcus pneu
6	47	48.0	326	22	AAW01092 CFE 95 protein seq
7	46	46.9	50	22	AAW84080 Human immune/haema
8	46	46.9	64	21	AAW57264 Human prostate can
9	46	46.9	123	19	AAW34009 Anti-CD4 antibody
10	45	45.9	285	23	AAW51648 Rat GPCR polypepti

11	45	45.9	298	15	AAW48755 Human thoracic aor
12	45	45.9	298	17	AAW02727 Human thoracic aor
13	45	45.9	323	23	AAU76411 Novel human G-prot
14	45	45.9	327	23	AAE14729 Human G-protein co
15	45	45.9	343	17	AAW96145 G protein coupled
16	45	45.9	343	17	AAW97222 Human G-protein co
17	45	45.9	343	22	AAW88477 Human membrane or
18	45	45.9	343	23	ABB08348 R7A-like G protein
19	45	45.9	1299	22	ABB70025 Drosophila melanog
20	44	44.9	127	22	AAO06630 Human polypeptide
21	44	44.9	151	22	AAO06175 Human polypeptide
22	44	44.9	1440	22	ABB62712 Drosophila melanog
23	43	43.9	114	16	AAW75430 Human thyroid stim
24	43	43.9	119	17	AAW97336 Humanised antibody
25	43	43.9	147	22	AAE03940 Human gene 43 enco
26	43	43.9	167	23	ABB89662 Human polypeptide
27	43	43.9	179	22	AAW83192 Human stromal cell
28	43	43.9	221	20	AAW8495 Human stomach can
29	43	43.9	221	21	AAW5793 Human secreted pro
30	43	43.9	221	21	AAW32203 Human receptor mol
31	43	43.9	221	22	AAE03982 Human gene 43 enco
32	43	43.9	221	22	AAW83191 Human stromal cell
33	43	43.9	221	22	AAW75379 Human secreted pro
34	43	43.9	248	21	AAW53433 Human colon cancer
35	43	43.9	285	23	ABB90191 Human polypeptide
36	43	43.9	475	22	AAW35910 Helicobacter pylor
37	43	43.9	519	19	AAW98444 H. pylori GHPO 534
38	43	43.9	519	22	AAW35795 Helicobacter pylor
39	43	43.9	805	23	ABB93752 Herbicidally activ
40	42	42.9	42	22	AAW87391 Human gene 50 enco
41	42	42.9	42	23	ABG5375 Human albumin fusi
42	42	42.9	125	18	AAW08501 C6 human sfv antib
43	42	42.9	125	18	AAW08503 C6 human sfv antib
44	42	42.9	125	18	AAW08504 C6 human sfv antib
45	42	42.9	125	18	AAW08505 C6 human sfv antib

ALIGNMENTS

RESULT 1

AAE13851

ID AAE13851 standard; peptide; 16 AA.

XX AAE13851;

AC AAE13851;

XX 26-FEB-2002 (first entry)

DT Human T cell epitope related to lung tumour-specific protein.

XX Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;

KW antitense-therapy; vaccine; immune response; lung cancer;

XX T cell epitope.

XX Homo sapiens.

OS OS

XX WO200172295-A2.

PN 04-OCT-2001.

XX 28-MAR-2001; 2001WO-US09991.

XX 29-MAR-2000; 2000US-0538037.

PR 05-JUN-2000; 2000US-0588937.

PR 18-AUG-2000; 2000US-0640878.

PR 22-SEP-2000; 2000US-2345178.

PR 01-NOV-2000; 2000US-0704512.

PR 14-DEC-2000; 2000US-0738973.

XX (CORI-) CORIXA CORP.

PA Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;

XX Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;

PI

XX WPI; 2001-639201/73.
 XX New human lung-specific polynucleotides and polypeptides for the
 PT diagnosis and treatment of disease e.g. lung cancer -
 XX
 XX Claim 2; Page 378; 378pp; English.
 XX
 CC The invention relates to isolated lung tumour-specific proteins and
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and
 CC their antigen-presenting cells are useful for stimulating and/or
 CC expanding T cells specific for a tumour protein, and for inhibiting
 CC the development of cancer. The invention also relates to a composition
 CC useful for stimulating an immune response, and for treating cancer. The
 CC lung tumour specific oligonucleotide is useful in gene therapy and for
 CC diagnosis, detection and treatment of lung cancer. The present sequence
 CC is human T cell epitope related to lung tumour-specific protein.
 XX
 SQ Sequence 16 AA;
 Query Match 100.0%; Score 98; DB 22; Length 16;
 Best Local Similarity 100.0%; Pred. No. 1.1e-08;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FOANCGIDFIIFWIFW 16
 Db 1 FOANCGIDFIIFWIFW 16
 RESULT 2
 AAY54085
 ID AAY54085 standard; Protein; 504 AA.
 XX
 AC AAY54085;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Enzyme EPS11 involved in exopolysaccharide biosynthesis.
 XX
 KW Exopolysaccharide; EPS; ESP enzyme; EPS1; EPS2; EPS3; EPS4; EPS5; EPS6;
 KW EPS7; EPS8; EPS9; EPS10; EPS11; Lactobacillus helveticus strain LH59;
 KW activated D-galactose pyranose; saccharide; beta-glycosyltransferase;
 KW undecaprenyl-phosphate-glycosyl-1-phosphate-transferase;
 KW alpha-glycosyltransferase; EPS polymerase; glycosyltransferase;
 KW phosphofuranose; transporter; food; fermented milk product; yoghurt;
 KW cheese; flavour stability; organoleptic property.
 XX
 OS Lactobacillus helveticus.
 XX
 PN WO9962316-A2.
 XX
 PD 09-DEC-1999.
 XX
 PF 22-APR-1999; 99WO-EP02841.
 XX
 PR 22-APR-1998; 98EP-0201310.
 PR 22-APR-1998; 98EP-0201311.
 PR 22-APR-1998; 98EP-0201312.
 XX
 PA (NEST) SOC PROD NESTLE SA.
 XX
 PI Stingele F, Germond JE, Lamothe G;
 XX
 DR WPI; 2000-097267/08.
 DR N-PSDB; AAZ45259, AAY54082, AAY54083, AAY54084; AAY54085.
 XX
 XX New recombinant enzymes for synthesis of exopolysaccharides,
 PT particularly in lactic acid bacteria, for improving properties of
 PT fermented milk products -
 XX
 XX Example 8; Page 130-131; 162pp; French.
 PS
 XX AAY54075-85 represent enzymes involved in the biosynthesis of
 CC

CC exopolysaccharides (EPS). These enzymes are designated EPS1-EPS11. and
 CC are encoded by open reading frames eps1-eps11. The proteins are isolated
 CC from Lactobacillus helveticus strain LH59. The proteins are used
 CC in a method for the synthesis of EPS, which includes at least one step
 CC of forming a bond (alpha or beta-isomer) between C-1 (carrying the
 CC reducing aldehyde function, of an activated D-galactose pyranose), and
 CC a phosphate on a lipophilic or proteinaceous primer. Stepwise synthesis
 CC of EPS occurs with, in each step, addition of a new sugar unit, through
 CC its hemi-acetyl function, to an alcoholic hydroxyl of a second sugar
 CC unit, present at the end of a chain of sugar residues bonded to the
 CC primer. EPS1 has homology with an undecaprenyl-phosphate-glycosyl-1-
 CC phosphate-transferase; EPS2, EPS3 and EPS4 are alpha-
 CC glycosyltransferases; EPS5 and EPS6 have homology with a beta-
 CC glycosyltransferase; EPS7 has homology with an EPS polymerase; EPS8
 CC is a glycosyltransferase; EPS9 catalyses the transfer of phosphofuranose
 CC onto the following repetitive unit; EPS10 transports the polysaccharides
 CC formed; and EPS11 is involved in the synthesis and export of formed
 CC polysaccharides. The EPS enzyme are used to improve properties of foods,
 CC particularly fermented milk products such as yoghurt and cheese,
 CC e.g. their organoleptic properties and flavour stability.
 XX
 SQ Sequence 504 AA;
 Query Match 51.0%; Score 50; DB 21; Length 504;
 Best Local Similarity 37.5%; Pred. No. 8.6;
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 QY 1 FOANCGIDFIIFWIFW 16
 Db 114 YRSKCIVTFMHFWFW 129
 RESULT 3
 AAY43787
 ID AAY43787 standard; Protein; 504 AA.
 XX
 AC AAY43787;
 XX
 DT 11-FEB-2000 (first entry)
 XX
 DE Amino acid sequence of the eps11 of Lactobacillus helveticus LH59.
 XX
 KW eps operon; Lactobacillus helveticus LH59; enzyme; eps1; eps2;
 KW eps3; eps4; eps5; eps6; eps7; eps8; eps9; eps10; eps11;
 KW exopolysaccharide biosynthesis; EPS; intersugar bond; antitumour;
 KW probiotic; foodstuff; organoleptic quality; flavour;
 KW lactic acid bacteria; acidified milk product; yoghurt; cheese.
 XX
 OS Lactobacillus helveticus.
 XX
 PN WO9954475-A2.
 XX
 PD 28-OCT-1999.
 XX
 PF 22-APR-1999; 99WO-EP03011.
 XX
 PR 22-APR-1998; 98EP-0201310.
 PR 22-APR-1998; 98EP-0201311.
 PR 22-APR-1998; 98EP-0201312.
 XX
 PA (NEST) SOC PROD NESTLE SA.
 XX
 PI Stingele F, Germond JE, Lamothe G;
 XX
 DR WPI; 2000-013255/01.
 DR N-PSDB; AAZ30356, AAY43784, AAY43785, AAY43786, AAY43787.
 XX
 XX New recombinant enzymes for biosynthesis of exopolysaccharides having
 PT e.g. antitumour or probiotic properties or useful in fermented milk
 PT products -
 XX
 XX Example 8; Page 131-132; 163pp; French.
 PS
 XX

CC AAY43777-87 represent enzymes encoded by the eps operon of *Lactobacillus*
 CC *helveticus* LH59. The operon contains 11 open reading frames, and
 CC encodes enzymes (eps1, eps2, eps3, eps4, eps5, eps6, eps7, eps8, eps9
 CC and eps11) that are involved in the biosynthesis of exopolysaccharides.
 CC (EPS). The enzymes catalyse the formation of specific intersugar bonds.
 CC The enzymes catalyse a process that includes at least one step of
 CC forming a bond (in alpha or beta anomeric form) between C1, carrying
 CC the reducing aldehyde group of an activated D-Galp (galactose in
 CC pyranose form), and a phosphate group on a lipophilic or proteinaceous
 CC primer. The enzymes are used to produce EPS that have antitumor or
 CC probiotic properties or are used in foodstuffs to improve organoleptic
 CC qualities and flavour. When expressed by lactic acid bacteria, EPS
 CC impart a free-flowing character and/or a smooth, creamy texture to
 CC acidified milk products (yoghurt or cheese).
 CC
 CC
 CC

SQ Sequence 504 AA;

Query Match 51.0%; Score 50; DB 21; Length 504;

Best Local Similarity 37.5%; Pred. No. 8.6;

Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 FOANCGIDFIIFWIFW 16

Db 114 YRSKCIYTFMHFWFFW 129

RESULT 4

AAW99067

ID AAW99067 standard; Protein; 207 AA.

AC AAW99067;

DT 14-MAY-1999 (first entry)

DE Streptococcus pneumoniae mray transferase ORF protein sequence.

XX Streptococcus pneumoniae; mray; transferase; infection; pneumonia;
 KW phospho-N-acetylmuramoyl-pentapeptide-transferase family; otitis media;
 KW bacteraemia; conjunctivitis; sinusitis; pleural empyema; endocarditis;
 KW meningitis; gene therapy.
 XX
 XX Streptococcus pneumoniae.
 XX
 XX EP897007-A2.
 XX
 XX 17-FEB-1999.
 XX
 XX 11-JUN-1998; 98EP-0304635.
 XX
 XX 16-APR-1998; 98US-0061156.
 XX
 XX 12-AUG-1997; 97US-0055467.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 XX (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 XX Fuyo J.L., Jaworski D.D., Kosmatka A.L., Lonetto M.A;
 XX Traini CM, Wang M;
 XX
 XX WPI: 1999-134240/12.
 XX N-PSDB; AAX18943.
 XX
 XX New Phospho-N-acetylmuramoyl-pentapeptide-transferase (mray)
 XX polypeptide and polynucleotide - useful as diagnostic reagents and
 XX for prevention and treatment of Streptococcus pneumoniae
 XX infections, which cause otitis media and meningitis
 XX
 XX Claim 1; Page 4; 37pp; English.
 XX
 XX The present sequence represents the ORF protein sequence of an isolated
 CC Phospho-N-acetylmuramoyl-pentapeptide-transferase designated mray. mray
 CC polynucleotides and proteins are useful for diagnosing susceptibility to
 CC diseases by detecting mutations or polymorphisms in the mray gene or
 CC analysing for the presence of amount of mray protein expressed in a

CC patient sample. mray PCR probes are useful for diagnosing diseases, and
 CC can characterise the response of the infectious organism to drugs. mray
 CC proteins and polynucleotides are also useful for screening for
 CC antagonists, agonists and drugs against infectious micro-organisms. mray
 CC agonists and antagonists are bacteriostatic and bacteriocidal compounds
 CC which can be used in treatment to enhance (agonist) or block (antagonist
 CC or antisense sequence) mray activity, therefore treating microbial
 CC diseases, especially Streptococcus pneumoniae diseases including otitis
 CC media, bacteraemia, conjunctivitis, pneumonia, sinusitis, pleural
 CC empyema, endocarditis and especially meningitis. Epitopes of mray
 CC proteins and polynucleotides are useful immunogens (vaccines) for
 CC producing anti-mray antibodies for prevention of bacterial infections,
 CC and mray polynucleotides can be used in genetic immunisation (gene
 CC therapy) to prevent infections. mray proteins, polynucleotides and their
 CC (antagonists) can prevent adhesion of bacteria to matrix proteins, and
 CC are useful for use on wounds and body implants to prevent bacterial
 CC infection.
 CC
 CC

SQ Sequence 207 AA;

Query Match 48.0%; Score 47; DB 20; Length 207;

Best Local Similarity 43.8%; Pred. No. 11;

Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 FOANCGIDFIIFWIFW 16

Db 24 YQVHLGIFYIVFALFW 39

RESULT 5

AAW99066

ID AAW99066 standard; Protein; 326 AA.

AC AAW99066;

DT 14-MAY-1999 (first entry)

DE Streptococcus pneumoniae mray transferase.

XX Streptococcus pneumoniae; mray; transferase; infection; pneumonia;
 KW phospho-N-acetylmuramoyl-pentapeptide-transferase family; otitis media;
 KW bacteraemia; conjunctivitis; sinusitis; pleural empyema; endocarditis;
 KW meningitis; gene therapy.
 XX
 XX Streptococcus pneumoniae.
 XX
 XX EP897007-A2.
 XX
 XX 17-FEB-1999.
 XX
 XX 11-JUN-1998; 98EP-0304635.
 XX
 XX 16-APR-1998; 98US-0061156.
 XX
 XX 12-AUG-1997; 97US-0055467.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 XX (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 XX Fuyo J.L., Jaworski D.D., Kosmatka A.L., Lonetto M.A;
 XX Traini CM, Wang M;
 XX
 XX WPI: 1999-134240/12.
 XX N-PSDB; AAX18942.
 XX
 XX New Phospho-N-acetylmuramoyl-pentapeptide-transferase (mray)
 XX polypeptide and polynucleotide - useful as diagnostic reagents and
 XX for prevention and treatment of Streptococcus pneumoniae
 XX infections, which cause otitis media and meningitis
 XX
 XX Claim 1; Page 4; 37pp; English.
 XX
 XX The present sequence represents an isolated Phospho-N-acetylmuramoyl-
 CC pentapeptide-transferase designated mray. mray polynucleotides and

XX AC AAM51648;
 XX DT 20-FEB-2002 (first entry)
 XX DE Rat GPCR polypeptide.
 XX KW Rat; G-protein coupled receptor; GPCR; gene therapy;
 KW MAS proto-oncogene receptor; human protease; disease.
 XX OS Rattus sp.
 XX PN WO200181409-A2.
 XX PD 01-NOV-2001.
 XX PF 24-APR-2001; 2001WO-US13097.
 XX PR 24-APR-2000; 2000US-199149P.
 XX PR 04-AUG-2000; 2000US-0633146.
 XX PA (PEKE) PE CORP NY.
 XX PI Wei M, Cravchik A, Di Francesco V, Beasley EM;
 XX WPI; 2002-049265/06.
 XX DR Novel human G protein-coupled receptor polypeptide that is related to
 PT MAS proto-oncogene receptor subfamily, useful as model and target for
 PT developing human therapeutic agent -
 XX PS Disclosure; Fig 2; 60pp; English.
 XX CC The invention relates to an isolated human G protein-coupled receptor
 CC (GPCR) polypeptide that is related to the MAS proto-oncogene receptor
 CC subfamily. The polypeptide comprises a fully defined sequence of 289
 CC amino acids as given in the specification, or its fragment comprising
 CC 10 contiguous amino acids, or an amino acid sequence of an allelic
 CC variant or orthologue of the amino acid sequence given in the sequence.
 CC The polypeptide is useful for identifying a modulator of a GPCR
 CC polypeptide or an agent that binds to it. The polypeptide is also
 CC useful for treating a disease or condition mediated by human proteases.
 CC The present sequence is a rat GPCR polypeptide used in comparison
 CC studies with the polypeptide of the invention.
 XX SQ Sequence 285 AA;
 Query Match 45.9%; Score 45; DB 23; Length 285;
 Best Local Similarity 60.0%; Pred. No. 30;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 6 GIDFIIFWIF 15
 Db 213 GIDWFLFWVF 222
 RESULT 11
 AAR48755
 ID AAR48755 standard; Protein; 298 AA.
 XX AC AAR48755;
 XX DT 07-JUN-1996 (first entry)
 XX DE Human thoracic aorta G-protein coupled receptor protein.
 XX KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
 KW psychotic disorder; schizophrenia; dopamine; cAMP; adenosine; thrombin;
 KW muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;
 KW rhodopsin; opsin; odorant; cytomegalovirus.
 XX OS Homo sapiens.
 XX PI Murphy RB, Schuster DI;
 XX PA (UUNY) UNIV NEW YORK STATE.

PN WO9405695-A1.
 XX 17-MAR-1994.
 XX PF 09-SEP-1993; 93WO-US08528.
 XX PR 10-SEP-1992; 92US-0943236.
 XX PA (UUNY) UNIV NEW YORK STATE.
 XX PI Murphy RB, Schuster DI;
 XX WPI; 1994-101120/12.
 XX DR Polypeptides of G-coupled receptor proteins (GPRs) - useful for
 PT binding GPR ligands or modulating GPR binding
 PT Disclosure; Page 130-131; 160pp; English.
 XX PS Proteins AAR48685-R48758 represent a range of G-protein coupled receptor
 CC proteins selected from cAMP, adenosine, muscarinic acetylcholine,
 CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
 CC odorant, cytomegalovirus and other G-protein coupled receptors. The
 CC receptor proteins were used to design polypeptides, pref. based on the
 CC transmembrane domains, for use in G-protein coupled receptor ligand
 CC binding assays. The polypeptide fragments retain biological activity
 CC such as binding a GPR ligand or modulating GPR ligand binding to a GPR
 CC (see AAR48759-R48758, AAR50569-R50807 and AAR89189-R89195 for examples
 CC of polypeptide fragments). The polypeptide fragments can be used in
 CC compositions for treating subjects suffering from a pathology related to
 CC a GPR abnormality e.g. a psychotic disorder such as schizophrenia.
 XX SQ Sequence 298 AA;
 Query Match 45.9%; Score 45; DB 15; Length 298;
 Best Local Similarity 60.0%; Pred. No. 31;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 6 GIDFIIFWIF 15
 Db 211 GIDWFLFWVF 220
 RESULT 12
 AAW02727
 ID AAW02727 standard; peptide; 298 AA.
 XX AC AAW02727;
 XX DT 13-NOV-1996 (first entry)
 XX DE Human thoracic aorta G-protein coupled receptor.
 XX KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
 KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
 KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
 KW odorant; cytomegalovirus; serotonergic.
 XX OS Homo sapiens.
 XX PN US5508384-A.
 XX PD 16-APR-1996.
 XX PF 10-SEP-1992; 92US-0943236.
 XX PR 09-SEP-1993; 93US-0118270.
 XX PR 10-SEP-1992; 92US-0943236.
 XX PA (UUNY) UNIV NEW YORK STATE.
 XX PI Murphy RB, Schuster DI;
 XX SQ

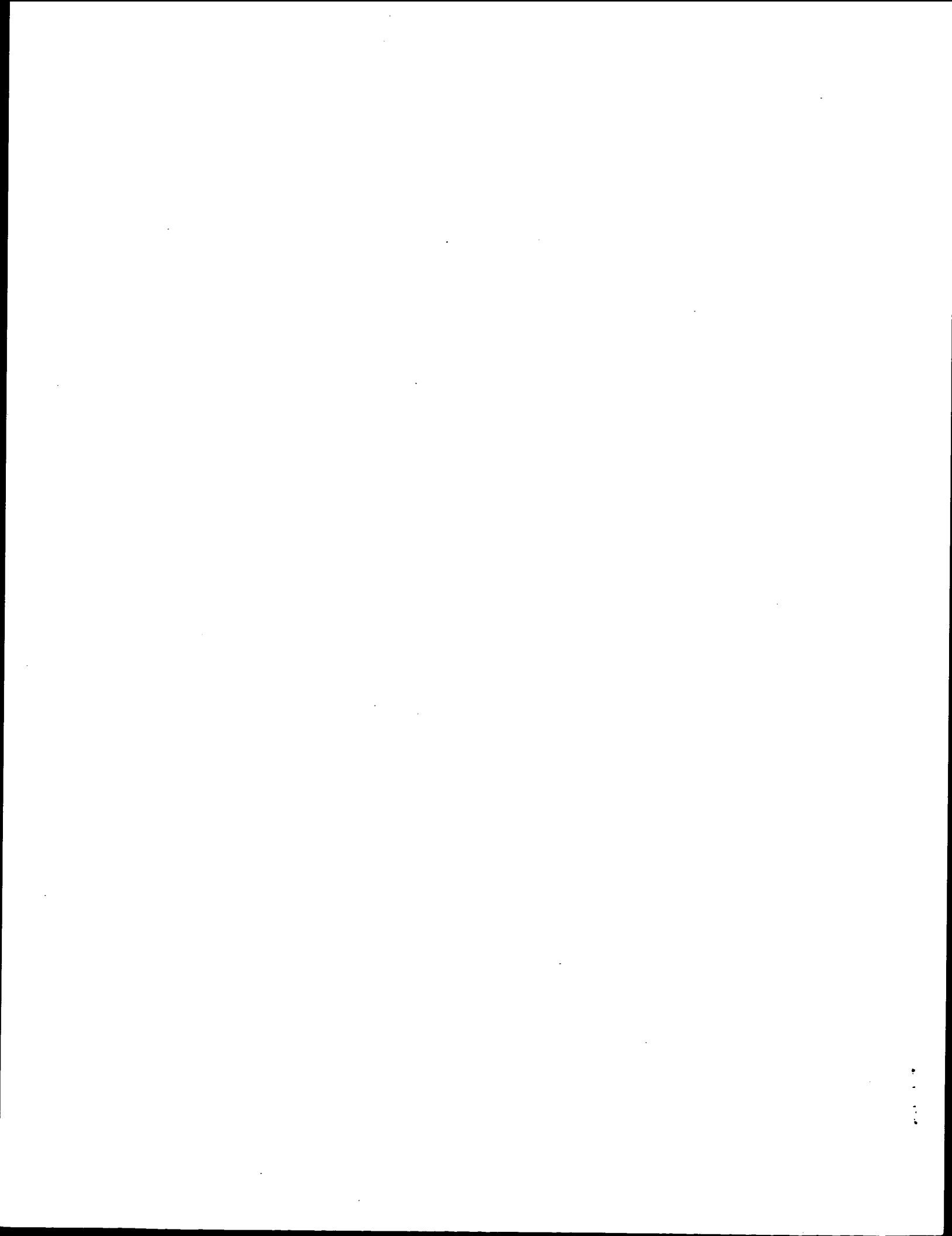
DR WPI; 1996-208785/21.
 PT New dopamine receptor peptide - useful as antipsychotic agent, e.g.
 PT for treating schizophrenia
 XX
 PS Disclosure; Column 173-176; 184pp; English.
 XX
 CC Proteins AA02657-W02730 represent a range of G-protein coupled receptor
 CC (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
 CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
 CC odorant, cytomegaloviral and other GPR proteins. The receptor proteins
 CC were used to design polypeptides, pref. based on the transmembrane
 CC domains, for use in G-protein coupled receptor ligand binding assays.
 CC The polypeptide fragments retain biological activity such as binding a
 CC GPR ligand or modulating GPR ligand binding to a GPR (see
 CC AA02747-W02999 for examples of polypeptide fragments). The polypeptide
 CC fragments can be used in compositions for treating subjects suffering
 CC from a pathology related to a GPR abnormality e.g. a psychotic disorder
 CC such as schizophrenia.
 XX
 SQ Sequence 298 AA;
 Query Match 45.9%; Score 45; DB 17; Length 298;
 Best Local Similarity 60.0%; Pred. No. 31;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 6 GIDFIIFWIF 15
 III: :||:
 Db 211 GIDWFLFWVF 220
 RESULT 13
 AAU76411
 ID AAU76411 standard; Protein; 323 AA.
 AC AAU76411;
 XX
 XX 08-MAY-2002 (first entry)
 DE Novel human G-protein coupled receptor IGPCR18.
 XX
 KW G-protein coupled; receptor; IGPCR18; cardiovascular system disorder;
 KW signal processing; female reproductive tissue; infertility;
 KW cardiovascular disease; coronary heart disease; heart attack; stroke;
 KW inflammatory disorder; metabolic disorder; uterus; placenta; ovary;
 KW prostate; reproductive disorder; pain; cancer; CNS disorder;
 KW central nervous system disorder; schizophrenia;
 KW episodic paroxysmal anxiety; EPA; obsessive compulsive disorder; COD;
 KW Parkinson's disease; multiple sclerosis; Alzheimer's disease; dementia;
 KW Huntington's disease; anorexia; kidney disease; renal failure; obesity;
 KW gastrointestinal disorder; irritable bowel syndrome; IBS; diarrhoea;
 KW motility disorder; gastric emptying; osteoporosis; infection; asthma;
 KW allergy; arthritis; sepsis.
 XX
 OS Homo sapiens.
 XX
 PN W0200202598-A2.
 XX
 XX 10-JAN-2002.
 PD
 XX 02-JUL-2001; 2001WO-EP07530.
 PF
 XX 30-JUN-2000; 2000US-215879P.
 PR
 XX (INGE-) INGENIUM PHARM AG.
 PA
 XX Wattler F, Wattler S, Trommler P, Nehls MC;
 PI
 XX WPI; 2002-140079/18.
 DR N-PSDB; ABK15133.
 DR
 XX New human G protein-coupled receptor protein, IGPCR18, useful for
 PT diagnosis, prevention, amelioration or treatment of pain, cancer,

PT inflammatory, metabolic, reproductive, gynaecological, and
 PT cardiovascular disorders
 XX
 PS Claim 8; Fig 2; 55pp; English.
 XX
 CC The invention describes a novel human G protein-coupled receptor (GPCR)
 CC protein, IGPCR18 (I). (I), the polynucleotide (II), agonists and
 CC antagonists of (I)/(II) are useful for the diagnosis and treatment of
 CC cardiovascular system disorders, the detection of mutant or
 CC inappropriately expressed forms of IGPCR18, and for drug screening.
 CC (I) is useful in treatment of diseases associated with signal processing
 CC in female reproductive tissues, such as infertility, cardiovascular
 CC diseases such as coronary heart disease, heart attack and stroke,
 CC inflammatory disorders and metabolic disorders linked to reproductive
 CC tissues like uterus, placenta, ovary and prostate, reproductive
 CC disorders, pain, cancer, central nervous system disorders such as
 CC schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as
 CC obsessive compulsive disorder (COD), Parkinson's disease, multiple
 CC sclerosis, Alzheimer's disease/dementia, Huntington's disease, anorexia,
 CC kidney diseases such as renal failure, obesity, gastrointestinal
 CC disorders such as irritable bowel syndrome (IBS), diarrhoea, motility
 CC disorders and conditions of delayed gastric emptying, osteoporosis,
 CC infections such as bacterial, fungal, protozoal and viral infections,
 CC asthma, allergy, arthritis, and sepsis. (I) is useful for the
 CC generation of antibodies, in identification of other cellular gene
 CC products involved in regulating IGPCR18, and as pharmaceutical reagents.
 CC (II) is useful for mapping the location of the gene to the chromosome,
 CC and to detect abnormalities involving IGPCR18 gene structure. This
 CC is the amino acid sequence of the novel human G-protein coupled receptor,
 CC IGPCR18, described in the method of the invention.
 XX
 SQ Sequence 323 AA;
 Query Match 45.9%; Score 45; DB 23; Length 323;
 Best Local Similarity 60.0%; Pred. No. 34;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 6 GIDFIIFWIF 15
 III: :||:
 Db 238 GIDWFLFWVF 247
 RESULT 14
 AAEL14729
 ID AAEL14729 standard; Protein; 327 AA.
 XX
 AC AAEL14729;
 XX
 XX 27-AUG-2002 (first entry)
 DT
 XX Human G-protein coupled receptor, IGS43.
 DE
 XX IGS43; G-protein coupled receptor; GPCR; uterus; lung; trachea;
 KW colon; small intestine; stomach; mammary gland; prostate; testis;
 KW psychiatric disorder; central nervous system disorder; schizophrenia;
 KW episodic and paroxysmal anxiety disorder; Parkinson's disease;
 KW multiple sclerosis; Alzheimer's disease; cardiovascular disease;
 KW heart failure; angina pectoris; kidney disease; dyslipidaemias; obesity;
 KW emesis; gastrointestinal disorder; inflammatory bowel disease; diabetes;
 KW osteoporosis; inflammation; infection; human immunodeficiency virus; HIV;
 KW cancer; immune disorder; urinary retention; asthma; allergy; arthritis;
 KW benign prostatic hypertrophy; endotoxin shock; sepsis; gene therapy;
 KW gynaecological disorder; vaccine; human.
 XX
 OS Homo sapiens.
 XX
 XX W0200228897-A2.
 PN
 XX 11-APR-2002.
 PD
 XX 28-SEP-2001; 2001WO-EP11319.
 PF
 XX 02-OCT-2000; 2000EP-0203411.
 PR

PR 04-OCT-2000; 2000US-237394P.
 XX (SOLV) SOLVAY PHARM BV.
 PA Deleersnijder W, Blockx H, De Moor L;
 PI WPI; 2002-426102/45.
 XX N-PSDB; AAD31159.
 DR Novel G-protein coupled receptor, termed IGS43 polypeptide and nucleic
 DR acid encoding the polypeptide, useful for treating disorders of uterus,
 DR kidney, lung, colon, stomach, mammary gland, prostate and testis .
 XX Claim 25; Page 7; 59pp; English.
 XX The invention relates to IGS43 G-protein coupled receptor (GPCR) and the
 CC polynucleotide encoding it. The IGS43 polypeptide, polynucleotide,
 CC agonist, antagonist or antibody is useful for treating dysfunctions or
 CC disorders related to uterus, kidney, lung, trachea, colon, small
 CC intestine, stomach, mammary gland, prostate, testis, central nervous
 CC system, cerebellum and spinal cord. The polypeptide is useful as vaccine
 CC for inducing immunological response in a mammal, for treating
 CC psychiatric and central nervous system disorders including
 CC schizophrenia, episodic and paroxysmal anxiety disorders e.g. obsessive
 CC compulsive disorder, bipolar disorder, Parkinson's disease, major
 CC anxiety disorder, autism, delirium, multiple sclerosis, Alzheimer's
 CC disease, dementia, severe mental retardation, Huntington's disease,
 CC dyskinesias, Tourette's syndrome, tics, tremor, dystonia, spasms,
 CC anorexia, bulimia, stroke, addiction/dependency/craving, sleep
 CC disorder, epilepsy, migraine, attention deficit/hyperactivity disorder,
 CC cardiovascular diseases including heart failure, angina pectoris,
 CC arrhythmias, myocardial infarction, cardiac hypertrophy, hypotension,
 CC hypertension, thrombosis, arteriosclerosis, cerebral vasospasm,
 CC subarachnoid haemorrhage, cerebral ischaemia, cerebral infarction,
 CC peripheral vascular disease, Raynaud's disease, kidney disease,
 CC dyslipidaemias, obesity, emesis, gastrointestinal disorders including
 CC irritable bowel syndrome, inflammatory bowel disease, diabetic
 CC gastroparesis and diabetes, ulcers, diarrhoea, osteoporosis,
 CC inflammations, infections including bacterial, fungal, protozoan and
 CC viral infections, particularly human immunodeficiency virus (HIV)-1 or
 CC HIV-2 infections, pain, cancers, chemotherapy induced injury, tumour
 CC invasion, immune disorders, urinary retention, asthma, allergies,
 CC arthritis, benign prostatic hypertrophy, endotoxin shock, sepsis,
 CC complications of diabetes mellitus, and gynaecological disorders.
 CC The polypeptide and polynucleotide of the invention are also useful as
 CC research reagents and materials for discovery of treatments and
 CC diagnostics to animal and human diseases. The polynucleotide is
 CC also useful for chromosome identification. The polypeptide is also useful
 CC for assessing the binding of small molecule substrates and ligands in
 CC cells, cell-free preparations, chemical libraries and natural product
 CC mixtures. The present sequence is human IGS43 GPCR.
 XX
 SQ Sequence 327 AA;
 Query Match 45.9%; Score 45; DB 23; Length 327;
 Best Local Similarity 60.0%; Pred. No. 34;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 6 GIDFIIFWIF 15
 DB 240 GIDWFLFWVF 249
 RESULT 15
 AAR96145
 ID AAR96145 standard; Protein; 343 AA.
 XX
 AC AAR96145;
 XX
 DT 29-OCT-1996 (first entry)
 XX
 DE G protein coupled receptor protein.

XX G protein coupled receptor protein; GCR; identification; detection;
 KW ligands; physiological response; arachidonic acid; acetylcholine;
 KW calcium; cyclic adenosine monophosphate; cAMP; inositol;
 KW membrane potential; phosphorylation; c-fos; antibodies;
 KW transgenic animals.
 XX
 OS Oryctolagus cuniculus.
 XX
 PN EP711831-A2.
 XX
 PD 15-MAY-1996.
 XX
 PF 11-NOV-1995; 95EP-0117786.
 XX
 PR 24-AUG-1995; 95JP-0215798.
 PR 14-NOV-1994; 94JP-0279545.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Fuji R, Hinuma S, Kawamata Y;
 XX
 XX WPI; 1996-232096/24.
 DR N-PSDB; AAT27218.
 DR
 XX New isolated rabbit G protein receptor protein - used partic. to
 PT identify agonists or antagonists which can promote or inhibit
 PT physiological responses
 XX
 PS Claim 1; Page 30-32; 40pp; English.
 XX
 XX G protein coupled receptor (GCR) proteins can be used to identify
 CC ligands which act as antagonists of GCR/ligand binding. Such ligands
 CC can be used for promoting or inhibiting physiological responses such
 CC as liberation of arachidonic acid, acetylcholine and endocellular
 CC calcium, endocellular cyclic AMP production, production of inositol,
 CC changes in cell membrane potential, phosphorylation of endocellular
 CC proteins, activation of c-fos, lowering of pH, activation of G
 CC protein and cell promulgation. The GCR products can also be used for
 CC the detection of ligands and in the production of antibodies and
 CC transgenic animals.
 XX
 SQ Sequence 343 AA;
 Query Match 45.9%; Score 45; DB 17; Length 343;
 Best Local Similarity 60.0%; Pred. No. 36;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 6 GIDFIIFWIF 15
 DB 256 GIDWFLFWVF 265

Search completed: July 9, 2003, 15:07:41
 Job time : 9.92035 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 4.95575 Seconds
(without alignments)
94.994 Million cell updates/sec

Title: US-09-854-133-587

Perfect score: 98

Sequence: 1 FOANGCIDFIIFWIFW 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	46	46.9	123	1	US-08-478-039-82
2	46	46.9	123	1	US-08-476-349A-82
3	45	45.9	298	1	US-08-118-270-76
4	45	45.9	298	5	PCT-US93-08528-76
5	42	42.9	125	2	US-08-665-202-45
6	42	42.9	125	2	US-08-665-202-46
7	42	42.9	125	2	US-08-665-202-49
8	42	42.9	125	2	US-08-665-202-51
9	42	42.9	342	2	US-08-483-151-2
10	42	42.9	518	4	US-09-134-001C-4069
11	41	41.8	119	2	US-08-318-157B-2
12	41	41.8	119	2	US-08-318-157B-12
13	41	41.8	119	2	US-08-318-157B-17
14	41	41.8	125	2	US-08-665-202-48
15	41	41.8	125	2	US-08-665-202-50
16	41	41.8	125	2	US-08-665-202-52
17	41	41.8	125	2	US-08-665-202-53
18	41	41.8	125	2	US-08-665-202-54
19	41	41.8	125	2	US-08-665-202-55
20	41	41.8	125	2	US-08-665-202-57
21	41	41.8	139	4	US-09-136-315-2
22	41	41.8	139	4	US-09-136-315-6
23	40	40.8	119	2	US-08-318-157B-8
24	40	40.8	119	2	US-08-318-157B-9
25	40	40.8	119	2	US-08-318-157B-10
26	40	40.8	119	2	US-08-318-157B-11
27	40	40.8	119	2	US-08-318-157B-13

Sequence 14, Appl
Sequence 15, Appl
Sequence 13, Appl
Sequence 21, Appl
Sequence 7, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 15, Appl
Sequence 23, Appl
Sequence 2, Appl
Sequence 74, Appl
Sequence 44, Appl
Sequence 10, Appl
Sequence 5, Appl
Sequence 3, Appl

28 40 40.8 119 2 US-08-318-157B-14
29 40 40.8 119 2 US-08-318-157B-15
30 40 40.8 203 4 US-09-124-141-13
31 40 40.8 203 4 US-09-124-141-21
32 40 40.8 593 4 US-08-592-696-4
33 40 40.8 594 2 US-09-027-536-4
34 40 40.8 594 3 US-09-028-148-4
35 40 40.8 594 4 US-09-124-141-15
36 40 40.8 594 4 US-09-124-141-23
37 40 40.8 625 1 US-08-242-689-2
38 40 40.8 625 4 US-08-468-583-2
39 40 40.8 1269 3 US-08-781-891-74
40 39.5 40.3 36 2 US-08-576-626A-44
41 39 39.8 114 3 US-08-483-749A-10
42 39 39.8 262 1 US-08-480-882B-5
43 39 39.8 262 1 US-08-480-210-5
44 39 39.8 262 1 US-08-220-401-3
45 39 39.8 262 1 US-08-220-401-3

ALIGNMENTS

RESULT 1

US-08-478-039-82
; Sequence 82, Application US/08478039
; Patent No. 5681722
; GENERAL INFORMATION:
; APPLICANT: Newman, Roland A.
; APPLICANT: Hanna, Nabil
; APPLICANT: Raab, Ronald W.
; TITLE OF INVENTION: Recombinant Antibodies for Human Therapy
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince St.
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,039
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/379,072
; FILING DATE: 25-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/912,292
; FILING DATE: 10-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/856,281
; FILING DATE: 23-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/735,064
; FILING DATE: 25-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin Esq., Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 012712-160
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 123 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant

;; TOPOLOGY: not relevant
;; MOLECULE TYPE: peptide
;; ORIGINAL SOURCE: Monkey
;; ORGANISM: Monkey
;; POSITION IN GENOME:
;; CHROMOSOME/SEGMENT: VH5 clone 5-11
US-08-478-039-82

Query Match 46.9%; Score 46; DB 1; Length 123;
Best Local Similarity 58.3%; Pred. No. 3.9;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 CGIDFIIFWIFW 16
|| | ||| |
Db 25 CGFSFTGFWSW 36

RESULT 2

US-08-476-349A-82
; Sequence 82, Application US/08476349A
; Patent No. 5750105
; GENERAL INFORMATION:
; APPLICANT: Newman, Roland A.
; APPLICANT: Hanna, Nabil
; APPLICANT: Raab, Ronald W.
; TITLE OF INVENTION: Recombinant Antibodies for Human Therapy
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince St.
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/476,349A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/379,072

FILING DATE: 25-JAN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/912,292

FILING DATE: 10-JUL-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/856,281

FILING DATE: 23-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/735,064

FILING DATE: 25-JUL-1991

ATTORNEY/AGENT INFORMATION:

NAME: Teskin Esq., Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 012712-161

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6620

TELEFAX: 703-836-2021

INFORMATION FOR SEQ ID NO: 82:

SEQUENCE CHARACTERISTICS:

LENGTH: 123 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: peptide

ORIGINAL SOURCE:

ORGANISM: Monkey

POSITION IN GENOME:

CHROMOSOME/SEGMENT: VH5 clone 5-11

US-08-476-349A-82

Query Match 46.9%; Score 46; DB 1; Length 123;
Best Local Similarity 58.3%; Pred. No. 3.9;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 CGIDFIIFWIFW 16
|| | ||| |
Db 25 CGFSFTGFWSW 36

RESULT 3

US-08-118-270-76
; Sequence 76, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
; APPLICANT: Murphy, Randall B.
; APPLICANT: Schuster, David I.
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/118,270

FILING DATE: 09-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/943,236

FILING DATE: 10-SEP-1992

ATTORNEY/AGENT INFORMATION:

NAME: Townsend, Kevin G.

REGISTRATION NUMBER: 34,033

REFERENCE/DOCKET NUMBER: MURPHY-2A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

TELEX: 248633

INFORMATION FOR SEQ ID NO: 76:

SEQUENCE CHARACTERISTICS:

LENGTH: 298 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-118-270-76

Query Match 45.9%; Score 45; DB 1; Length 298;
Best Local Similarity 60.0%; Pred. No. 14;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15
||| :||: |
Db 211 GIDWFLFWVF 220

RESULT 4

PCT-US93-08528-76
; Sequence 76, Application PC/TUS9308528
; GENERAL INFORMATION:
; APPLICANT: New York University
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: BROWDY AND NEIMARK
;; STREET: 419 Seventh Street, N.W., Suite 300
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20004
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US93/08528
;; FILING DATE: 09-SEP-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/943,236
;; FILING DATE: 10-SEP-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Townsend, Kevin G.
;; REGISTRATION NUMBER: 34,033
;; REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-628-5197
;; TELEFAX: 202-737-3528
;; TELEX: 248633
;; INFORMATION FOR SEQ ID NO: 76:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 298 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
PCT-US93-08528-76

Query Match 45.9%; Score 45; DB 5; Length 298;
Best Local Similarity 60.0%; Pred. No. 14;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFFWIF 15
|||: |||:
Db 211 GIDWFLFWVF 220

RESULT 5
US-08-665-202-45
; Sequence 45, Application US/08665202
; Patent No. 5977322
; GENERAL INFORMATION:
; APPLICANT: Marks, James D.
; APPLICANT: Schlier, Robert
; TITLE OF INVENTION: No. 5977322el High Affinity Human Antibodies to
; TUMOR ANTIGENS
; NUMBER OF SEQUENCES: 141
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,202
; FILING DATE: 13-JUN-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/000,238
; FILING DATE: 15-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 02307E-061410
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 125 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/000,250
;; FILING DATE: 15-JUN-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hunter, Tom
;; REGISTRATION NUMBER: 38,498
;; REFERENCE/DOCKET NUMBER: 02307E-061410
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 576-0200
;; TELEFAX: (415) 576-0300
;; INFORMATION FOR SEQ ID NO: 45:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 125 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-665-202-45

Query Match 42.9%; Score 42; DB 2; Length 125;
Best Local Similarity 54.5%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 GIDFIIFFWIF 16
|||: |||:
Db 26 GYDFTTYWIAW 36

RESULT 6
US-08-665-202-46
; Sequence 46, Application US/08665202
; Patent No. 5977322
; GENERAL INFORMATION:
; APPLICANT: Marks, James D.
; APPLICANT: Schlier, Robert
; TITLE OF INVENTION: No. 5977322el High Affinity Human Antibodies to
; TUMOR ANTIGENS
; NUMBER OF SEQUENCES: 141
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,202
; FILING DATE: 13-JUN-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/000,238
; FILING DATE: 14-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 02307E-061410
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 125 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

; MOLECULE TYPE: peptide
US-08-665-202-46

Query Match 42.9%; Score 42; DB 2; Length 125;
Best Local Similarity 54.5%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 6 GIDFTIFWIFW 16
| | | : | | |
Db 26 GYDFTTYWIAW 36

RESULT 7

US-08-665-202-49
; Sequence 49, Application US/08665202
; Patent No. 5977322

; GENERAL INFORMATION:

; APPLICANT: Marks, James D.

; APPLICANT: Schier, Robert

; TITLE OF INVENTION: No. 5977322el High Affinity Human Antibodies to

; TITLE OF INVENTION: Tumor Antigens

; NUMBER OF SEQUENCES: 141

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/665,202

; FILING DATE: 13-JUN-1996

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/000,238

; FILING DATE: 14-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/000,250

; FILING DATE: 15-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Hunter, Tom

; REGISTRATION NUMBER: 38,498

; REFERENCE/DOCKET NUMBER: 02307E-061410

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 49:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 125 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-665-202-49

Query Match

Best Local Similarity 42.9%; Score 42; DB 2; Length 125;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 6 GIDFTIFWIFW 16
| | | : | | |
Db 26 GYDFTTYWIAW 36

RESULT 8

US-08-665-202-51

; Sequence 51, Application US/08665202
; Patent No. 5977322

; GENERAL INFORMATION:

; APPLICANT: Marks, James D.

; APPLICANT: Schier, Robert

; TITLE OF INVENTION: No. 5977322el High Affinity Human Antibodies to

; TITLE OF INVENTION: Tumor Antigens

; NUMBER OF SEQUENCES: 141

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/665,202

; FILING DATE: 13-JUN-1996

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/000,238

; FILING DATE: 14-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/000,250

; FILING DATE: 15-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Hunter, Tom

; REGISTRATION NUMBER: 38,498

; REFERENCE/DOCKET NUMBER: 02307E-061410

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 51:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 125 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-665-202-51

Query Match

Best Local Similarity 42.9%; Score 42; DB 2; Length 125;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 6 GIDFTIFWIFW 16
| | | : | | |
Db 26 GYDFTTYWIAW 36

RESULT 9

US-08-483-151-2

; Sequence 2, Application US/08483151

; Patent No. 5858752

; GENERAL INFORMATION:

; APPLICANT: Seed, Brian

; APPLICANT: Holgersson, Jan

; TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,151
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Lech, Karen F.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/278001
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-151-2

Query Match 42.9%; Score 42; DB 2; Length 342;
Best Local Similarity 45.5%; Pred. No. 45;
Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 6 GIDFIIIFW 16
| | | | |
Db 21 GATFWIIFW 31

RESULT 10

US-09-134-001C-4069
Sequence 4069, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4069

LENGTH: 518
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
FEATURE:
NAME/KEY: UNSURE
LOCATION: (6)
OTHER INFORMATION: Identity of amino acid at the above locations are unknown.

US-09-134-001C-4069

Query Match 42.9%; Score 42; DB 4; Length 518;
Best Local Similarity 58.3%; Pred. No. 69;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 QANGGIDFIIFW 13
: | | | | |
Db 175 RSCGIGPKGEW 186

RESULT 11

US-08-318-157B-2
Sequence 2, Application US/08318157B
Patent No. 5874540
GENERAL INFORMATION:

APPLICANT: HANSEN, Hans J.
APPLICANT: ARMOUR, Kathryn L.
TITLE OF INVENTION: CDR-GRAFTED TYPE III ANTI-CEA HUMANIZED

TITLE OF INVENTION: MOUSE MONOCLONAL ANTIBODIES
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318,157B
FILING DATE: 05-OCT-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: SAXE, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 18733/464
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 119 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-318-157B-2

Query Match 41.8%; Score 41; DB 2; Length 119;
Best Local Similarity 42.9%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 3 ANCGIDFIIFW 16
| | | | |
Db 23 AASGDFTTYMSW 36

RESULT 12

US-08-318-157B-12
Sequence 12, Application US/08318157B
Patent No. 5874540
GENERAL INFORMATION:

APPLICANT: HANSEN, Hans J.
APPLICANT: ARMOUR, Kathryn L.
TITLE OF INVENTION: CDR-GRAFTED TYPE III ANTI-CEA HUMANIZED
TITLE OF INVENTION: MOUSE MONOCLONAL ANTIBODIES
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318,157B
FILING DATE: 05-OCT-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: SAXE, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 18733/464

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 119 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-318-157B-12

Query Match 41.8%; Score 41; DB 2; Length 119;
Best Local Similarity 35.7%; Pred. No. 21;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 ANCGIDFIIFWIFW 16
: : : : : :
Db 23 SSSGDFDTTYMWSW 36

RESULT 13

US-08-318-157B-17
Sequence 17, Application US/08318157B
Patent No. 5874540

GENERAL INFORMATION:

APPLICANT: HANSEN, Hans J.
APPLICANT: ARMOUR, Kathryn L.
TITLE OF INVENTION: CDR-GRAFTED TYPE III ANTI-CRA HUMANIZED
TITLE OF INVENTION: MOUSE MONOCLONAL ANTIBODIES
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318,157B
FILING DATE: 05-OCT-1994

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: SAXE, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 18733/464
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:
LENGTH: 119 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-318-157B-17

Query Match 41.8%; Score 41; DB 2; Length 119;
Best Local Similarity 35.7%; Pred. No. 21;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 ANCGIDFIIFWIFW 16
: : : : : :
Db 23 SSSGDFDTTYMWSW 36

RESULT 14

US-08-665-202-48
Sequence 48, Application US/08665202
Patent No. 5977322

GENERAL INFORMATION:

APPLICANT: Marks, James D.
APPLICANT: Schier, Robert
TITLE OF INVENTION: No. 5977322el High Affinity Human Antibodies to
TITLE OF INVENTION: Tumor Antigens
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/665,202

FILING DATE: 13-JUN-1996

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/000,238

FILING DATE: 14-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/000,250

FILING DATE: 15-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Hunter, Tom

REGISTRATION NUMBER: 38,498

REFERENCE/DOCKET NUMBER: 02307E-061410

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 48:

SEQUENCE CHARACTERISTICS:

LENGTH: 125 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-665-202-48

Query Match

41.8%; Score 41; DB 2; Length 125;

Best Local Similarity 54.5%; Pred. No. 22;

Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 6 GIDFIFWIFW 16

: : : : : :
Db 26 GIDFSTYWIW 36

RESULT 15

US-08-665-202-50

Sequence 50, Application US/08665202

Patent No. 5977322

GENERAL INFORMATION:

APPLICANT: Marks, James D.

APPLICANT: Schier, Robert

TITLE OF INVENTION: No. 5977322el High Affinity Human Antibodies to

TITLE OF INVENTION: Tumor Antigens

NUMBER OF SEQUENCES: 141

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

;;
;; ZIP: 94111-3834
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/665,202
;; FILING DATE: 13-JUN-1996
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/000,238
;; FILING DATE: 14-JUN-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/000,250
;; FILING DATE: 15-JUN-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hunter, Tom
;; REGISTRATION NUMBER: 38,498
;; REFERENCE/DOCKET NUMBER: 02307E-061410
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 576-0200
;; TELEFAX: (415) 576-0300
;; INFORMATION FOR SEQ ID NO: 50:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 125 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-665-202-50

Query Match 41.8%; Score 41; DB 2; Length 125;
Best Local Similarity 54.5%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 GIDFIIFWIFW 16
Db 26 GIDFSYWIW 36

Search completed: July 9, 2003, 15:08:27
Job time : 5.95575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:06:38 ; Search time 5.38053 Seconds
(without alignments)
346.196 Million cell updates/sec

Title: US-09-854-133-587

Perfect score: 98
Sequence: 1 FOANGGIDFIIFWIFW 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	98	100.0	16	9	US-09-854-133-587
2	98	100.0	16	9	US-10-144-649A-587
3	98	100.0	16	10	US-09-738-973-587
4	98	100.0	97	9	US-09-854-133-586
5	98	100.0	97	9	US-10-144-649A-586
6	98	100.0	97	10	US-09-738-973-586
7	98	100.0	114	9	US-10-144-649A-742
8	46	46.9	64	10	US-09-925-300-1842
9	46	46.9	123	10	US-09-850-165-90
10	45	45.9	323	9	US-10-266-643-4
11	45	45.9	323	10	US-09-816-087-4
12	45	45.9	343	9	US-10-176-079-2
13	45	45.9	343	9	US-10-176-079-9
14	45	45.9	343	9	US-09-929-752-2
15	45	45.9	343	9	US-09-929-752-9
16	45	45.9	343	9	US-10-225-567A-482
17	45	45.9	343	9	US-10-184-426-2
18	45	45.9	343	9	US-10-184-426-9
19	45	45.9	343	10	US-09-985-694A-2

```
20 45 45.9 343 10 US-09-985-694A-9 Sequence 9, Appli
21 43 43.9 248 9 US-09-925-299-973 Sequence 973, App
22 43 43.9 248 10 US-09-925-299-973 Sequence 973, App
23 43 43.9 475 10 US-09-815-242-11503 Sequence 11503, A
24 43 43.9 519 9 US-09-895-913A-118 Sequence 118, App
25 43 43.9 519 10 US-09-815-242-11388 Sequence 11388, A
26 42.5 43.4 393 10 US-09-789-561-132 Sequence 2, Appli
27 42 42.9 42 10 US-09-789-561-132 Sequence 132, App
28 42 42.9 619 10 US-09-729-094-4 Sequence 4, Appli
29 41 41.8 77 9 US-10-091-504-1073 Sequence 1073, Ap
30 41 41.8 77 10 US-09-764-869-1073 Sequence 1073, Ap
31 41 41.8 119 10 US-09-253-794-2 Sequence 2, Appli
32 41 41.8 119 10 US-09-253-794-12 Sequence 12, Appli
33 41 41.8 119 10 US-09-253-794-17 Sequence 17, Appli
34 41 41.8 230 9 US-09-738-626-5746 Sequence 5746, Ap
35 41 41.8 310 9 US-09-804-291-485 Sequence 485, App
36 41 41.8 310 9 US-10-183-116-51 Sequence 51, Appli
37 41 41.8 310 10 US-09-886-055-485 Sequence 485, App
38 41 41.8 676 9 US-10-295-403-100 Sequence 100, App
39 40 40.8 71 10 US-09-925-302-881 Sequence 881, App
40 40 40.8 119 10 US-09-253-794-8 Sequence 8, Appli
41 40 40.8 119 10 US-09-253-794-9 Sequence 9, Appli
42 40 40.8 119 10 US-09-253-794-10 Sequence 10, Appli
43 40 40.8 119 10 US-09-253-794-11 Sequence 11, Appli
44 40 40.8 119 10 US-09-253-794-13 Sequence 13, Appli
45 40 40.8 119 10 US-09-253-794-14 Sequence 14, Appli
```

ALIGNMENTS

```
RESULT 1
US-09-854-133-587
; Sequence 587, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 587
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-133-587
```

```
Query Match 100.0%; Score 98; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.5e+08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 FOANGGIDFIIFWIFW 16
    |||||
DB 1 FOANGGIDFIIFWIFW 16
```

```
RESULT 2
US-10-144-649A-587
; Sequence 587, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Rongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
```

```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144,649A
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 587
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-649A-587

Query Match      100.0%; Score 98; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FOANCGIDFIIFWIFW 16
       |||||
Db      1 FOANCGIDFIIFWIFW 16
       |||||

RESULT 3
US-09-738-973-587
; Sequence 587, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSEQ for Windows Version 3.0.
; SEQ ID NO 587
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-587
```

```
Query Match      100.0%; Score 98; DB 10; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FOANCGIDFIIFWIFW 16
       |||||
Db      1 FOANCGIDFIIFWIFW 16
       |||||

RESULT 4
US-09-854-133-586
; Sequence 586, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
```

```
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 586
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-133-586

Query Match      100.0%; Score 98; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. 7.8e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FOANCGIDFIIFWIFW 16
       |||||
Db      35 FOANCGIDFIIFWIFW 50
       |||||

RESULT 5
US-10-144-649A-586
; Sequence 586, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Lique
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144,649A
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 586
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-649A-586

Query Match      100.0%; Score 98; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. 7.8e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FOANCGIDFIIFWIFW 16
       |||||
Db      35 FOANCGIDFIIFWIFW 50
       |||||

RESULT 6
US-09-738-973-586
; Sequence 586, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
```

; CURRENT FILLING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 586
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-586

Query Match 100.0%; Score 98; DB 10; Length 97;
Best Local Similarity 100.0%; Pred. No. 7.8e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANGCIDFIIFW 16
|||
DB 35 FOANGCIDFIIFW 50

RESULT 7
US-10-144-649A-742
; Sequence 742, Application US/10144649A
; Publication No. US2003011859A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144,649A
; CURRENT FILLING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 742
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-649A-742

Query Match 100.0%; Score 98; DB 9; Length 114;
Best Local Similarity 100.0%; Pred. No. 9e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANGCIDFIIFW 16
|||
DB 52 FOANGCIDFIIFW 67

RESULT 8
US-09-925-300-1842
; Sequence 1842, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILLING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILLING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILLING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1842
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1842

Query Match 46.9%; Score 46; DB 10; Length 64;
Best Local Similarity 50.0%; Pred. No. 4;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 QANGCIDFIIFW 13
::|||::|:
DB 29 ESNGLDFVSFF 40

RESULT 9
US-09-850-165-90
; Sequence 90, Application US/09850165
; Patent No. US20020150580A1
; GENERAL INFORMATION:
; APPLICANT: NEWMAN, ROLAND A.
; APPLICANT: HANNA, NABIL
; APPLICANT: RAAB, RONALD W.
; TITLE OF INVENTION: RECOMBINANT ANTIBODIES FOR HUMAN THERAPY
; FILE REFERENCE: 037003-0280614
; CURRENT APPLICATION NUMBER: US/09/850,165
; CURRENT FILLING DATE: 2001-05-08
; PRIOR APPLICATION NUMBER: 09/082,472
; PRIOR FILLING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: 08/476,237
; PRIOR FILLING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/397,072
; PRIOR FILLING DATE: 1995-04-17
; PRIOR APPLICATION NUMBER: 07/912,292
; PRIOR FILLING DATE: 1992-07-10
; PRIOR APPLICATION NUMBER: 07/856,281
; PRIOR FILLING DATE: 1992-03-23
; PRIOR APPLICATION NUMBER: 07/735,064
; PRIOR FILLING DATE: 1991-07-25
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 90
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: monkey clone
US-09-850-165-90

Query Match 46.9%; Score 46; DB 10; Length 123;
Best Local Similarity 58.3%; Pred. No. 7.3;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 CGIDFIIFW 16
|||
DB 25 CGFSFTGFWISW 36

RESULT 10
US-10-266-643-4
; Sequence 4, Application US/10266643
; Publication No. US20030059891A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000749-CON
; CURRENT APPLICATION NUMBER: US/10/266,643
; CURRENT FILLING DATE: 2002-10-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-266-643-4

Query Match 45.9%; Score 45; DB 9; Length 323;
Best Local Similarity 60.0%; Pred. No. 25;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15
|||: |||:|
Db 244 GIDWFLFWVF 253

RESULT 11

US-09-816-087-4
; Sequence 4, Application US/09816087
; Patent No. US20020064822A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; FILE REFERENCE: CLO00749-CIP
; CURRENT APPLICATION NUMBER: US/09/816,087
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-816-087-4

Query Match 45.9%; Score 45; DB 10; Length 323;
Best Local Similarity 60.0%; Pred. No. 25;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15
|||: |||:|
Db 244 GIDWFLFWVF 253

RESULT 12

US-10-176-079-2
; Sequence 2, Application US/10176079
; Publication No. US20020192760A1
; GENERAL INFORMATION:
; APPLICANT: Li et al.
; TITLE OF INVENTION: G-Protein Coupled Receptor
; FILE REFERENCE: PF145PIDIC1
; CURRENT APPLICATION NUMBER: US/10/176,079
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: 09/562,909
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: 08/461,989
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: PCT/US94/13296
; PRIOR FILING DATE: 1994-11-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 343
; TYPE: PRT
; ORGANISM: human
US-10-176-079-2

Query Match 45.9%; Score 45; DB 9; Length 343;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15
|||: |||:|
Db 256 GIDWFLFWVF 265

RESULT 13

US-10-176-079-9

; Sequence 9, Application US/10176079
; Publication No. US20020192760A1
; GENERAL INFORMATION:
; APPLICANT: Li et al.

; TITLE OF INVENTION: G-Protein Coupled Receptor
; FILE REFERENCE: PF145PIDIC1
; CURRENT APPLICATION NUMBER: US/10/176,079
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: 09/562,909
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: 08/461,989
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: PCT/US94/13296
; PRIOR FILING DATE: 1994-11-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 343
; TYPE: PRT
; ORGANISM: human
US-10-176-079-9

Query Match 45.9%; Score 45; DB 9; Length 343;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15
|||: |||:|
Db 256 GIDWFLFWVF 265

RESULT 14

US-09-929-752-2
; Sequence 2, Application US/09929752
; Publication No. US20030113909A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; Fujii, Ryo
; Kawamata, Yuji
; TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,
; PRODUCTION AND USE THEREOF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; Zip: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/929,752
; FILING DATE: 14-Aug-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/555,905
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 7-215798
; FILING DATE: 24-AUG-1995
; APPLICATION NUMBER: JP 6-326611
; FILING DATE: 28-DEC-1994
; APPLICATION NUMBER: JP 7-007177
; FILING DATE: 20-JAN-1995
; APPLICATION NUMBER: JP 7-057186
; FILING DATE: 16-MAR-1995
; APPLICATION NUMBER: JP 7-224544
; FILING DATE: 10-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Conlin, David G.
; REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 1550/45836
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 343 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-929-752-2

Query Match 45.9%; Score 45; DB 9; Length 343;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15
|||: :||:|
Db 256 GIDWFLFWVF 265

RESULT 15

US-09-929-752-9
Sequence 9, Application US/09929752
Publication No. US20030113909A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
Fujii, Ryo
Kawamata, Yuji
TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,
PRODUCTION AND USE THEREOF
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/929,752
FILING DATE: 14-Aug-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/555,905
FILING DATE: <Unknown>
APPLICATION NUMBER: JP 7-215798
FILING DATE: 24-AUG-1995
APPLICATION NUMBER: JP 6-326611
FILING DATE: 28-DEC-1994
APPLICATION NUMBER: JP 7-007177
FILING DATE: 20-JAN-1995
APPLICATION NUMBER: JP 7-057186
FILING DATE: 16-MAR-1995
APPLICATION NUMBER: JP 7-224544
FILING DATE: 10-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 1550/45836
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 343 amino acids

TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-929-752-9

Query Match 45.9%; Score 45; DB 9; Length 343;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15
|||: :||:|
Db 256 GIDWFLFWVF 265

Search completed: July 9, 2003, 15:10:29
Job time : 6.38053 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 7.22124 Seconds
(without alignments)
213.004 Million cell updates/sec

Title: US-09-854-133-587
Perfect score: 98
Sequence: 1 FOANCGIDFIIFWIFW 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	51.0	544	2 S61985	ALG6 protein - yea
2	48	49.0	262	1 J01724	E1 membrane glycop
3	47	48.0	491	2 D95366	NADH2 dehydrogenas
4	45	45.9	343	2 A35639	G protein-coupled
5	44	44.9	179	1 D69890	conserved hypother
6	44	44.9	391	2 B96994	Fe-S oxidoreductas
7	43	43.9	284	2 AH0415	probable membrane-
8	43	43.9	332	2 B81252	NADH2 dehydrogenas
9	43	43.9	475	2 A71937	amino acid permeas
10	43	43.9	519	1 A64647	amino acid permeas
11	43	43.9	545	2 B89823	hypothetical prote
12	43	43.9	804	1 Y0M0	sucrose synthase (
13	42	42.9	121	2 A30560	Ig heavy chain v r
14	42	42.9	147	2 S64252	probable membrane
15	42	42.9	194	2 AB2320	hypothetical prote
16	42	42.9	256	2 T05554	hypothetical prote
17	42	42.9	329	1 C69516	hypothetical prote
18	42	42.9	483	2 S41689	cytochrome-c oxida
19	41	41.8	123	2 PH1423	Ig heavy chain v r
20	41	41.8	127	2 PH1409	Ig heavy chain v r
21	41	41.8	130	2 S06817	Ig heavy chain v r
22	41	41.8	134	2 PH1422	Ig heavy chain v r
23	41	41.8	194	2 PH1071	Ig heavy chain v r
24	41	41.8	216	2 E75425	ATPase subunit 6 -
25	41	41.8	342	2 E64395	hypothetical prote
26	41	41.8	345	2 T25580	malic acid transpo
27	41	41.8	495	1 S25942	hypothetical prote
28	41	41.8	585	2 T19814	NADH2 dehydrogenas
29	40.5	41.3	123	2 E71163	hypothetical prote

30	40.5	41.3	318	2 AD2556	hypothetical prote
31	40.5	41.3	666	2 S59109	NADH2 dehydrogenas
32	40.5	41.3	838	2 T47828	hypothetical prote
33	40.5	41.3	1037	2 E84723	hypothetical prote
34	40	40.8	70	2 B82692	hypothetical prote
35	40	40.8	97	2 T03811	hypothetical prote
36	40	40.8	119	2 S36257	Ig heavy chain v r
37	40	40.8	221	2 T28778	hypothetical prote
38	40	40.8	289	2 G96981	ABC-type sugar tra
39	40	40.8	309	2 AC0489	probable membrane
40	40	40.8	337	2 S47742	hypothetical 37.9k
41	40	40.8	337	2 B91179	hypothetical prote
42	40	40.8	337	2 C86025	hypothetical prote
43	40	40.8	342	2 AE0986	probable membrane
44	40	40.8	375	2 C64216	hypothetical prote
45	40	40.8	378	2 G86220	hypothetical prote

ALIGNMENTS

RESULT 1

S61985
N:ALG6 protein - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 20-Jun-2000
C:Accession: S61985; S66867; S72134
R:Sterky, F.; Uhlen, M.
submitted to the EMBL Data Library, December 1995
A:Reference number: S61981
A:Accession: S61985
A:Molecule type: DNA
A:Residues: 1-544 <STE>
A:Cross-references: EMBL:U43491; NID:g1150992; PIDN:AAC49481.1; PID:g1150997
R:Pettersson, B.; Sterky, F.; Uhlen, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66682
A:Accession: S66687
A:Molecule type: DNA
A:Residues: 1-544 <PET>
A:Cross-references: EMBL:274910; NID:g1420089; PIDN:CAA99190.1; PID:g1420090; MIPS:
A:Experimental source: strain S288C
R:Sterky, F.; Holmberg, A.; Pettersson, B.; Uhlen, M.
Yeast 12, 1091-1095, 1996
A:Title: The sequence of a 30 kb fragment on the left arm of chromosome XV from Sac
A:Reference number: S72130; MUID:97051599; PMID:8896276
A:Accession: S72134
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-544 <STW>
A:Cross-references: EMBL:U43491; NID:g1150992; PIDN:AAC49481.1; PID:g1150997
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1995
C:Genetics:
A:Gene: SGD:ALG6
A:Cross-references: SGD:S0005528; MIPS:YOR002W
A:Map position: 15R
C:Keywords: transmembrane protein
F:40-56/Domain: transmembrane #status predicted <TM1>
F:145-161/Domain: transmembrane #status predicted <TM2>
F:221-237/Domain: transmembrane #status predicted <TM3>
F:262-278/Domain: transmembrane #status predicted <TM4>
F:336-352/Domain: transmembrane #status predicted <TM5>
F:358-374/Domain: transmembrane #status predicted <TM6>
F:509-525/Domain: transmembrane #status predicted <TM7>

Query Match 51.0%; Score 50; DB 2; Length 544;
Best Local Similarity 61.5%; Pred. No. 3.8;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 4 NCGIDFIIFWIFW 16

DB 514 NCAVGICFISIFW 526

RESULT 2
 JQ1724
 E1 membrane glycoprotein precursor - canine coronavirus (strain Insavc-1)
 N:Alternate names: matrix glycoprotein
 C:Species: canine coronavirus
 C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 16-Jun-2000
 C:Accession: JQ1724
 R:Horsburgh, B.C.; Brierley, I.; Brown, T.D.K.
 J. Gen. Virol. 73, 2849-2862, 1992
 A:Title: Analysis of a 9.6 kb sequence from the 3' end of canine coronavirus genomic RNA
 A:Reference number: PQ0481; MUID:93057357; PMID:1431811
 A:Accession: JQ1724
 A:Molecule type: genomic RNA
 A:Residues: 1-262 <ROR>
 A:Cross-references: DDBJ:DI3096; NID:g406193; PIDN:BA02413.1; PID:g406200
 C:Genetics:
 A:Gene: M
 A:Superfamily: coronavirus E1 membrane glycoprotein
 C:Keywords: glycoprotein; matrix protein; transmembrane protein
 F:1-17/Domain: signal sequence #status predicted <SIG>
 F:18-262/Product: E1 membrane glycoprotein #status predicted <MAP>
 F:56-72/Domain: transmembrane #status predicted <TM1>
 F:85-101/Domain: transmembrane #status predicted <TM2>
 F:113-134/Domain: transmembrane #status predicted <TM3>
 F:32.55/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:33.57/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:44/Binding site: carbohydrate (Thr) (covalent) #status predicted

Query Match 49.0%; Score 48; DB 1; Length 262;
 Best Local Similarity 50.0%; Pred. No. 4;
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 CGIDFIIFWIF 16
 ||| : | : |
 DB 81 CGIKMLIMWLLW 92

RESULT 3
 D95366
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) Nuom2 [imported] - Sinorhizobium meliloti
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 03-Jun-2002
 C:Accession: D95366
 R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
 ; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
 A:Reference number: A95262; MUID:21396509; PMID:11481432
 A:Accession: D95366
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-491 <ROR>
 A:Cross-references: GB:AE006469; PIDN:AAK65494.1; PID:g14523966; GSPDB:GN00165
 A:Experimental source: strain 1021, megaplasmid pSymA
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hymen, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebaul, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: nuom2
 A:Genome: plasmid
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 4
 C:Keywords: oxidoreductase

Query Match 48.0%; Score 47; DB 2; Length 491;
 Best Local Similarity 41.7%; Pred. No. 10;

Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 5 CGIDFIIFWIF 16
 | : | : | : |
 DB 127 CALDLFLFYVFW 138

RESULT 4
 A35639
 G protein-coupled receptor RTA - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 16-Jul-1999
 C:Accession: A35639
 R:Ross, P.C.; Figler, R.A.; Corjay, M.H.; Barber, C.M.; Adam, N.; Marcus, D.R.; Ly
 Proc. Natl. Acad. Sci. U.S.A. 87, 3052-3056, 1990
 A:Title: RTA, a candidate G protein-coupled receptor: cloning, sequencing, and tis
 A:Reference number: A35639; MUID:90222168; PMID:2109324
 A:Accession: A35639
 A:Molecule type: mRNA
 A:Residues: 1-343 <ROR>
 A:Cross-references: GB:M35297; NID:g206809; PIDN:AAA42087.1; PID:g206810; GB:M32099
 C:Superfamily: mas transforming protein
 C:Keywords: G protein-coupled receptor; glycoprotein; membrane protein
 F:4/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 45.9%; Score 45; DB 2; Length 343;
 Best Local Similarity 60.0%; Pred. No. 15;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15
 ||| : | : |
 DB 256 GIDWFLFWVF 265

RESULT 5
 D69890
 conserved hypothetical protein yndM - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: D69890
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.;
 C.: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.
 A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrar
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.;
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulle
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lard
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Port
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Sc
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.;
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uch
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, E.; Yoshikawa, H.; Yumoto, K.; Yata, K.; Yosh
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Yoshikawa, H.; Yumoto, K.; Yata, K.; Yosh
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus sub
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: D69890
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-179 <ROR>
 A:Cross-references: GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CAB13667.1; PID:g262
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yndM
 C:Superfamily: hypothetical protein ycbP

Query Match 44.9%; Score 44; DB 1; Length 179;
 Best Local Similarity 46.2%; Pred. No. 12;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 ANCGIDFIIFWIF 15
 | : | : | : |
 DB 94 ADFGLSFVILWVF 106

A; Re:
A; Acc:
A; Sta:

C:Genetics:
A:Gene: all4113
C:Superfamily: Synechocystis hypothetical protein sill656
Query Match 42.9%; Score 42; DB 2; Length 194;
Best Local Similarity 37.5%; Pred No. 27;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
QY 1 FOANGCIDFIIFWIFW 16
Db 70 FWVCGVIALLFVYW 85

Search completed: July 9, 2003, 15:09:49
Job time : 12.2212 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 1.9823 seconds
(without alignments)
334.773 Million cell updates/sec

Title: US-09-854-133-587
Perfect score: 98
Sequence: 1 FQANGCIDFIIFWIFW 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	50	51.0	544	1 ALG6_YEAST	Q12001 saccharomyc
2	48	49.0	262	1 VME1_CVCAI	P36299 canine ente
3	45	45.9	343	1 RTA_RAT	P23749 rattus norv
4	45	45.9	406	1 G64B_DROME	P83294 drosophila
5	43	43.9	545	1 ARAB_STAAM	Q99w57 staphylococ
6	43	43.9	804	1 SUS2_ARATH	Q00917 arabidopsis
7	42	42.9	147	1 YGYO_YEAST	P53074 saccharomyc
8	42	42.9	329	1 YL31_ARCFU	O28149 archaeoglob
9	42	42.9	389	1 FUT7_MOUSE	Q11131 mus musculu
10	41	41.8	342	1 Y762_METJA	Q58172 methanococc
11	41	41.8	495	1 NU4M_MARPO	P26848 marchantia
12	40.5	41.3	666	1 NU5M_CHOCR	P48920 chondrus cr
13	40	40.8	337	1 YHJD_ECOLI	P37642 escherichia
14	40	40.8	375	1 Y147_MYCGE	P47393 mycoplasma
15	40	40.8	593	1 DCE1_RAT	P18088 rattus norv
16	40	40.8	594	1 DCE1_FELCA	P14748 felis silve
17	40	40.8	594	1 DCE1_HUMAN	Q59259 homo sapien
18	40	40.8	594	1 DCE1_PIG	P48319 sus scrofa
19	39.5	40.3	107	1 ELBA_ECOLI	P75987 escherichia
20	39	39.8	262	1 VGL1_CVPR8	P33464 porcine res
21	39	39.8	262	1 VME1_CVFFS	P09175 porcine tra
22	39	39.8	262	1 VME1_CVPPU	P04135 porcine tra
23	39	39.8	262	1 VME1_CVPRM	P24412 porcine res
24	39	39.8	262	1 VME1_FIPV	P25878 feline infe
25	39	39.8	310	1 RCE1_CHLAU	P11695 chloroflexu
26	39	39.8	329	1 MRAY_LACLA	Q9ch70 lactococcus
27	39	39.8	440	1 YA85_MYCPN	P75608 mycoplasma
28	39	39.8	679	1 MADE_TREPA	O83759 treponema p
29	38	38.8	116	1 HV36_MOUSE	P01806 mus musculu
30	38	38.8	117	1 HV42_MOUSE	P01812 mus musculu
31	38	38.8	118	1 HV39_MOUSE	P01809 mus musculu
32	38	38.8	119	1 HV37_MOUSE	P01807 mus musculu
33	38	38.8	119	1 HV38_MOUSE	P01808 mus musculu

34	38	38.8	119	1 HV40_MOUSE	P01810 mus musculu
35	38	38.8	236	1 CD63_BOVIN	Q9xsk2 bos taurus
36	38	38.8	245	1 TATC_CAMJE	Q9pht8 campylobact
37	38	38.8	282	1 AOP6_HUMAN	Q13520 homo sapien
38	38	38.8	283	1 NADC_METJA	Q57916 methanococc
39	38	38.8	317	1 YKG8_CAEEL	P46558 caenorhabdi
40	38	38.8	325	1 MCSR_MOUSE	P41149 mus musculu
41	38	38.8	325	1 MCSR_RAT	P35345 rattus norv
42	38	38.8	336	1 MRAY_STRPY	Q99YK2 streptococc
43	38	38.8	377	1 Y147_MYCPN	P75585 mycoplasma
44	38	38.8	398	1 Y432_MYCGE	Q49432 mycoplasma
45	38	38.8	404	1 Y432_MYCPN	P75166 mycoplasma

ALIGNMENTS

RESULT 1
ALG6_YEAST
ID ALG6_YEAST STANDARD; PRT; 544 AA.
AC Q12001;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dolichyl pyrophosphate Man9GlcNAc2 alpha-1,3-glucosyltransferase (EC 2.4.1.-) (Dolichyl-P-Glc:Man9GlcNAc2-P-dolichyl glucosyltransferase).
GN ALG6 OR YOR002W OR UNA544.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97051599; PubMed=8896276;
RA Sterky F., Holmberg A., Pettersson B., Uhlen M.;
RT "The sequence of a 30 kb fragment on the left arm of chromosome XV from Saccharomyces cerevisiae reveals 15 open reading frames, five of which correspond to previously identified genes.";
RL Yeast 12:1091-1095(1996).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=97031444; PubMed=8877369;
RA Reiss G., te Heesen S., Zimmerman J., Robbins P.W., Aebl M.;
RT "Isolation of the ALG6 locus of Saccharomyces cerevisiae required for glucosylation in the N-linked glucosylation pathway.";
RL Glycobiology 6:493-498(1996).
CC -!- FUNCTION: Adds the first glucose residue to the lipid-linked oligosaccharide precursor for N-linked glucosylation. Transfers glucose from dolichyl phosphate glucose (Dol-P-Glc) onto the lipid-linked oligosaccharide Man(9)GlcNAc(2)-pp-Dol.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum (Potential).
CC -!- SIMILARITY: BELONGS TO THE ALG6/ALG8 GLUCOSYLTRANSFERASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U43491; AAC49481.1; -;
CC EMBL; Z74910; CAA99190.1; -;
CC SGD; S0005528; ALG6.
CC InterPro; IPR004856; Alg6_Alg8.
CC Pfam; PF03155; Alg6_Alg8; 1.
CC Transferrase; Glycosyltransferase; Transmembrane;
CC Endoplasmic reticulum.
CC TRANSMEM 36 56
CC POTENTIAL.
CC TRANSMEM 105 125
CC POTENTIAL.
CC TRANSMEM 146 166

FT TRANSMEM 172 192 POTENTIAL.
 FT TRANSMEM 224 244 POTENTIAL.
 FT TRANSMEM 264 284 POTENTIAL.
 FT TRANSMEM 333 353 POTENTIAL.
 FT TRANSMEM 357 377 POTENTIAL.
 FT TRANSMEM 399 419 POTENTIAL.
 FT TRANSMEM 426 446 POTENTIAL.
 FT TRANSMEM 482 502 POTENTIAL.
 FT TRANSMEM 509 529 POTENTIAL.
 SQ SEQUENCE 544 AA; 62782 MW; 64BFA1A1F6D02B7 CRC64;

Query Match 51.08; Score 50; DB 1; Length 544;
 Best Local Similarity 61.5%; Pred. No. 1.3;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 NCIGDIFIFWIFW 16
 DB 514 NCAVGFCFSIFW 526

RESULT 2
 VME1_CVCAI
 ID VME1_CVCAI STANDARD; PRT; 262 AA.

AC P36299;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE El glycoprotein precursor (Matrix glycoprotein) (Membrane glycoprotein).
 GN M.

OS Canine enteric coronavirus (strain Insavc-1) (CCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Coronaviridae; Coronavirus.
 OX NCBI_TaxID=36391;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=93057357; PubMed=1431811;
 RA Horsburgh B.C., Brierley I., Brown T.D.K.;
 RT "Analysis of a 9.6 kb sequence from the 3' end of canine coronavirus genomic RNA."
 RL J. Gen. Virol. 73:2849-2862(1992).

CC -!- FUNCTION: TARGETS THE SITE OF VIRUS MORPHOGENESIS AND MAY BE IMPLICATED IN VIRAL PATHOGENESIS.
 CC -!- SUBCELLULAR LOCATION: LARGELY EMBEDDED IN THE LIPID BILAYER.
 CC -!- SIMILARITY: BELONGS TO THE CORONAVIRUS M PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC EMBL; D13096; BAA02413.1; -
 DR PIR; J01724; J01724.
 DR InterPro; IPR002574; Corona_M.
 DR Pfam; PF01635; Corona_M; 1.
 KW Matrix protein; Transmembrane; Glycoprotein; Signal.

FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 262 E1 GLYCOPROTEIN.
 FT TRANSMEM 56 72 POTENTIAL.
 FT TRANSMEM 85 101 POTENTIAL.
 FT TRANSMEM 115 134 POTENTIAL.
 FT CARBOHYD 32 32 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 33 33 O-LINKED (POTENTIAL).
 FT CARBOHYD 44 44 O-LINKED (POTENTIAL).
 FT CARBOHYD 55 55 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 262 AA; 23508 MW; 35C6D45DC6A474F6 CRC64;

Query Match 49.0%; Score 48; DB 1; Length 262;
 Best Local Similarity 50.0%; Pred. No. 1.4;
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 CGIDFIFWIFW 16
 DB 81 CGIKMLIMWLLW 92

RESULT 3

RTA_RAT
 ID RTA_RAT STANDARD; PRT; 343 AA.
 AC P23749;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-JUL-1993 (Rel. 26, Last annotation update)
 DE Probable G protein-coupled receptor RTA.
 GN RTA.

OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]

SEQUENCE FROM N.A.

RP STRAIN=Sprague-Dawley; TISSUE=Aorta;
 RX MEDLINE=90222168; PubMed=2109324;

RA ROSS P.C., Figler R.A., Corjay M.H., Barber C.M., Adam N.,
 RA Marcus D.R., Lynch K.K.;

RT "RTA, a candidate G protein-coupled receptor: cloning, sequencing, and tissue distribution."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3052-3056(1990).
 CC -!- FUNCTION: ORPHAN RECEPTOR.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: GUT, VAS DEFERENS, UTERUS, AND AORTA BUT ONLY BARELY DETECTABLE IN LIVER, KIDNEY, LUNG, AND SALIVARY GLAND. IN THE BRAIN, RTA IS MARKEDLY ABUNDANT IN THE CEREBELLUM.

CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS. MOST SIMILAR TO MAS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC EMBL; M35297; AAA42087.1; -
 DR EMBL; M35298; AAA42088.1; -
 DR PIR; A35639; A35639.

DR InterPro; IPR000276; GPCR_Rhodpsn.
 DR Pfam; PF00001; 7tm_1; 1.

DR PROSITE; PS00337; G-PROTEIN_RECEP_F1_1; 1.
 DR PROSITE; PS0262; G-PROTEIN_RECEP_F1_2; 1.

KW G-protein coupled receptor; Transmembrane; Glycoprotein.
 FT DOMAIN 1 44 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 45 66 1 (POTENTIAL).
 FT DOMAIN 67 82 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 83 104 2 (POTENTIAL).
 FT DOMAIN 105 123 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 124 144 3 (POTENTIAL).
 FT DOMAIN 145 160 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 161 181 4 (POTENTIAL).
 FT DOMAIN 182 198 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 199 220 5 (POTENTIAL).
 FT DOMAIN 221 241 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 242 263 6 (POTENTIAL).
 FT DOMAIN 264 273 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 274 294 7 (POTENTIAL).
 FT DOMAIN 295 343 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 4 4 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 343 AA; 38364 MW; E4630007770941F4 CRC64;

Query Match 45.9%; Score 45; DB 1; Length 343;
 Best Local Similarity 60.0%; Pred. No. 5.2;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Robertson H.,
Unpublished observations (NOV-2001).

CC -|- CATALYTIC ACTIVITY: ATP + L-ribose = ADP + L-ribose 5-phosphate.
 CC -|- PATHWAY: L-arabinose catabolism; second step.
 CC -|- SIMILARITY: BELONGS TO THE RIBULOSE KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF003359; BAB56714.1; -
 CC EMBL: AF003130; BAB41741.1; -
 CC InterPro: IPR000577; FGGY_kin.
 CC Pfam: PF00370; FGGY; 1.
 CC Pfam: PF02782; FGGY_C; 1.
 CC Transferase, Kinase; Arabinose catabolism; Complete proteome.
 KW SEQUENCE 545 AA; 60965 MW; E3917036237CEA4B CRC64;
 CC -----
 CC Query Match 43.9%; Score 43; DB 1; Length 545;
 CC Best Local Similarity 50.0%; Pred. No. 16;
 CC Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 CC -----
 CC QY 2 QANCGDIFIFW 13
 CC :|||:|
 CC 195 RSNCGLGFAFW 206
 CC -----
 CC RESULT 6
 CC SUS2_ARATH STANDARD; PRT; 804 AA.
 CC ID SUS2_ARATH
 CC AC Q00917;
 CC DT 01-APR-1993 (Rel. 25, Created)
 CC DT 01-APR-1993 (Rel. 25, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Sucrose synthase (EC 2.4.1.13) (sucrose-UDP glucosyltransferase).
 CC OS Arabidopsis thaliana (Mouse-ear cress).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 CC OX NCBI_TaxID=3702;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=cv. Columbia;
 CC RA MEDLINE=92119221; PubMed=1531031;
 CC RA Chopra S., Del-Favero J., Dolferus R., Jacobs M.;
 CC RT "Sucrose synthase of Arabidopsis: genomic cloning and sequence
 CC characterization".
 CC RL Plant Mol. Biol. 18:131-134(1992).
 CC CC -|- FUNCTION: SUCROSE-CLEAVING ENZYME THAT PROVIDES UDP-GLUCOSE AND
 CC FRUCTOSE FOR VARIOUS METABOLIC PATHWAYS.
 CC CC -|- CATALYTIC ACTIVITY: UDP-glucose + D-fructose = UDP + sucrose.
 CC CC -|- INDUCTION: BY ANAEROBIC STRESS.
 CC CC -|- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1. PLANT
 CC SUCROSE SYNTHASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X60987; CAA43303.1; -
 CC PIR: S19125; YUMU.
 CC InterPro: IPR001296; Glycosyl_transf_1.
 CC DR InterPro: IPR000368; Sucrose_synth.
 CC DR Pfam: PF00534; Glycosyl_transf_1; 1.
 CC DR Pfam: PF00862; Sucrose_synth; 1.
 CC KW Transferase; Glycosyltransferase; Multigene family.

SQ SEQUENCE 804 AA; 91989 MW; 3E727D3CDFF9A4B9 CRC64;
 CC -----
 CC Query Match 43.9%; Score 43; DB 1; Length 804;
 CC Best Local Similarity 62.5%; Pred. No. 23;
 CC Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 CC -----
 CC QY 9 PIIIFWIFW 16
 CC :|||:|
 CC 19 FLSEFWFV 26
 CC -----
 CC RESULT 7
 CC YGY0_YEAST STANDARD; PRT; 147 AA.
 CC ID YGY0_YEAST
 CC AC P53074;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 01-OCT-1996 (Rel. 34, Last annotation update)
 CC DE Hypothetical 17.3 kDa protein in SEC15-SAP4 intergenic region.
 CC GN YGL230C.
 CC OS Saccharomyces cerevisiae (Baker's yeast).
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CC OX NCBI_TaxID=4932;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RA Farmann B., Kramer B., Kramer W.;
 CC RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z72752; CAA96948.1; -
 CC DR SGD; S0003199; YGL230C.
 CC KW Hypothetical protein; Transmembrane.
 CC FT TRANSMEM 85 105 POTENTIAL.
 CC SQ SEQUENCE 147 AA; 17262 MW; F8F040A1D618CD96 CRC64;
 CC -----
 CC Query Match 42.9%; Score 42; DB 1; Length 147;
 CC Best Local Similarity 50.0%; Pred. No. 7.2;
 CC Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -----
 CC QY 5 CGDIFIFWIFW 16
 CC :|||:|
 CC 96 CAIQFLFIYW 107
 CC -----
 CC RESULT 8
 CC YL31_ARCFU STANDARD; PRT; 329 AA.
 CC ID YL31_ARCFU
 CC AC O28149;
 CC DT 16-OCT-2001 (Rel. 40, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Hypothetical protein Ar2131.
 CC GN Ar2131.
 CC OS Archaeoglobus fulgidus.
 CC OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 CC OC Archaeoglobaceae; Archaeoglobus.
 CC OX NCBI_TaxID=2234;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 CC RX MEDLINE=98049343; PubMed=9389475;
 CC RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 CC RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 CC RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
 CC RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
 RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon *Archaeoglobus fulgidus*.";
 RL Nature 390:364-370(1997).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF000957; AAB89128.1; -;
 DR TIGR; AF2131; -;
 DR
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 5 24 POTENTIAL.
 FT TRANSMEM 34 56 POTENTIAL.
 FT TRANSMEM 92 114 POTENTIAL.
 FT TRANSMEM 124 146 POTENTIAL.
 FT TRANSMEM 159 181 POTENTIAL.
 FT TRANSMEM 196 218 POTENTIAL.
 FT TRANSMEM 231 253 POTENTIAL.
 FT TRANSMEM 263 285 POTENTIAL.
 FT TRANSMEM 306 328 POTENTIAL.
 SQ SEQUENCE 329 AA; 37066 MW; C07CAFD427DC26F7 CRC64;
 Query Match 42.9%; Score 42; DB 1; Length 329;
 Best Local Similarity 71.4%; Pred. No. 15;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 10 IIFWIFW 16
 I::I::I
 DB 304 ILFWLFW 310
 RESULT 9
 FUT7_MOUSE
 ID FUT7_MOUSE STANDARD; PRT; 389 AA.
 AC Q11131;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Alpha-(1,3)-fucosyltransferase (EC 2.4.1.-) (Galactoside 3-L-
 DE fucosyltransferase) (Fucosyltransferase 7) (FUCT-VII).
 GN FUT7
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NIH Swiss;
 RX MEDLINE=96215226; PubMed=8626519;
 RA Smith P.L., Gersten K.M., Petryniak B., Kelly R.J., Rogers C.,
 RA Natsuka Y., Alford J.A. III, Scheidegger E.P., Natsuka S., Lowe J.B.;
 RT "Expression of the alpha(1,3)fucosyltransferase Fuc-TVII in lymphoid
 RT aggregate high endothelial venules correlates with expression of L-
 RT selectin ligands.";
 RL J. Biol. Chem. 271:8250-8259(1996).
 CC -!- FUNCTION: MAY CATALYSE ALPHA-1,3 GLYCOSIDIC LINKAGES INVOLVED IN
 CC THE EXPRESSION OF SIALYL LEWIS X ANTIGENS.
 CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + alpha-2,3-Neu-N-acetyl-1,4-
 CC beta-D-galactosyl-N-acetyl-D-glucosaminyl-R = GDP + alpha-2,3-
 CC Neu-N-acetyl-1,4-beta-D-galactosyl-(alpha-1,3-L-fucosyl)-
 CC N-acetyl-D-glucosaminyl-R.

CC -!- PATHWAY: Glycosylation.
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
 CC FORM IN TRANS CISTERNAE OF GOLGI.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG AND BONE MARROW AND
 CC TO A MUCH LESSER EXTENT IN SPLEEN, SALIVARY GLAND AND SKELETAL
 CC MUSCLE.
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 10.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U45980; AAC52484.1; -;
 DR MGD; MGI:107692; Fut7.
 DR InterPro: IPR001503; GT_10.
 DR Pfam: PF00852; Glyco_Transf_10; 1.
 KW Transferase; Glycosyltransferase; Transmembrane; Glycoprotein;
 KW Signal-anchor; Golgi stack; Alternative splicing.
 FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 56 78 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 79 389 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1 51 MPTCPGAGLSTPTGTHRLFPDPKAPSWESRKEATCNSSS
 FT PGPWAAPTQV -> MNCI (IN ISOFORM 2).
 SQ SEQUENCE 389 AA; 44494 MW; 118FC6B2378B99C6 CRC64;
 Query Match 42.9%; Score 42; DB 1; Length 389;
 Best Local Similarity 45.5%; Pred. No. 17;
 Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 6 GIDFIIFWIFW 16
 I::I::I::I
 DB 68 GATFMVWIFW 78
 RESULT 10
 Y762_METJA
 ID Y762_METJA STANDARD; PRT; 342 AA.
 AC Q58172;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0762.
 GN MJ0762.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, *Methanococcus*
 RT *jannaschii*.";
 RL Science 273:1058-1073(1996).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

```
CC -1- SIMILARITY: STRONG, TO M.JANNASCHII MJ0576 AND TO S.POMBE MALATE
CC PERMEASE (WAE1).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67521; AAB98753.1; -.
CC TIGR; MJ0762; -.
CC InterPro; IPR004695; C4dic_mal_transp.
CC Pfam; PF03595; C4dic_mal_tran; 1.
CC TIGRFAMs; TIGR00816; tdt; 1.
CC Hypothetical protein; Transmembrane; Transport; Complete proteome.
CC FT TRANSMEM 8 28 POTENTIAL.
CC TRANSMEM 39 59 POTENTIAL.
CC TRANSMEM 79 99 POTENTIAL.
CC TRANSMEM 108 128 POTENTIAL.
CC TRANSMEM 142 162 POTENTIAL.
CC TRANSMEM 175 195 POTENTIAL.
CC TRANSMEM 207 227 POTENTIAL.
CC TRANSMEM 242 262 POTENTIAL.
CC TRANSMEM 276 296 POTENTIAL.
CC TRANSMEM 304 324 POTENTIAL.
CC SEQUENCE 342 AA; 39534 MW; 08EFEC3E2C4955D8 CRC64;

Query Match 41.8%; Score 41; DB 1; Length 342;
Best Local Similarity 45.0%; Pred. No. 22;
Matches 9; Conservative 1; Mismatches 0; Indels 10; Gaps 1;

QY 6 GIDFI-----IFWIF 15
Db ||||| |||:|
93 GIDFILKNLFLGKIFWVF 112

RESULT 11
NU4M_MARPO
ID NU4M_MARPO STANDARD; PRT; 495 AA.
AC P26848;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3).
GN ND4 OR NAD4.
OS Marchantia polymorpha (Liverwort).
OC Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiopsida; Marchantiidae; Marchantiales; Marchantiaceae;
OC NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114051; PubMed=1731062;
RA Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N.,
RA Akashi K., Kanegae T., Ogura Y., Kohchi T., Ohyama K.;
RT "Gene organization deduced from the complete sequence of liverwort
RT Marchantia polymorpha mitochondrial DNA. A primitive form of plant
RT mitochondrial genome."
RL J. Mol. Biol. 223:1-7(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93247547; PubMed=8483448;
RA Nozato N., Oda K., Yamato K., Ohta E., Takemura M., Akashi K.,
RA Fukuzawa H., Ohyama K.;
RT "Cotranscriptional expression of mitochondrial genes for subunits of
RT NADH dehydrogenase, nad5, nad4, nad2, in Marchantia polymorpha."
RL Mol. Gen. Genet. 237:343-350(1993).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M68929; AAC09398.1; -.
CC PIR; S25942; S25942.
CC InterPro; IPR003918; NADHub_oxred4.
CC InterPro; IPR001750; Oxidored_q1.
CC Pfam; PF00361; oxidored_q1; 1.
CC PRINTS; PR01437; NUOXDRDTASE4.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
CC SEQUENCE 495 AA; 56311 MW; 0F75894D6CAAED4 CRC64;

Query Match 41.8%; Score 41; DB 1; Length 495;
Best Local Similarity 41.7%; Pred. No. 31;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 5 CGIDFIIFWIF 16
Db |||:|:|:|
135 CSLDLLIFYVFF 146

RESULT 12
NU5M_CHOCR
ID NU5M_CHOCR STANDARD; PRT; 666 AA.
AC P48920;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3).
GN ND5 OR NAD5.
OS Chondrus crispus (Carrageen).
OC Mitochondrion.
OC Eukaryota; Rhodophyta; Florideophyceae; Gigartinales; Gigartineae;
OC Chondrus.
OC NCBI_TaxID=2769;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Apices;
RX MEDLINE=95341681; PubMed=7616569;
RA Leblanc C., Boyen C., Richard O., Bonnard G., Grienenberger J.M.,
RA Kioareg B.;
RT "Complete sequence of the mitochondrial DNA of the rhodophyte
RT Chondrus crispus (Gigartinales). Gene content and genome
RT organization."
RL J. Mol. Biol. 250:484-495(1995).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z47547; CAA87625.1; -.
CC InterPro; IPR003916; NADHub_oxred5.
CC InterPro; IPR001750; Oxidored_q1.
CC InterPro; IPR001516; Oxidored_q1_N.
CC Pfam; PF00361; oxidored_q1; 1.
CC Pfam; PF00662; oxidored_q1_N; 1.
CC PRINTS; PR01434; NADHDHGNASE5.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
CC SEQUENCE 666 AA; 75628 MW; 37E86F2C24B9D360 CRC64;

Query Match 41.3%; Score 40.5; DB 1; Length 666;
Best Local Similarity 53.3%; Pred. No. 48;
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;
```

```

QY      1 FOANGGIDFIIFWIF 15
Db      176 FGLSLGI-FLIFWIF 189

RESULT 13
YHJD_ECOLI
ID YHJD_ECOLI STANDARD; PRT; 337 AA.
AC P37642;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yhjD.
GN YHJD OR B3522.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -!- SIMILARITY: STRONG, TO E.CHRYSANTHEMI HYPOTHETICAL PROTEIN IN
CC KDKG 5'REGION (AC P45417).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00039; AAC18498.1; -.
DR EMBL; AF000428; AAC76547.1; -.
DR EcoGene; EGI2248; yhjD.
DR InterPro; IPR005274; Cons_hypoth766.
DR Pfam; PF03631; Ribonuclease_BN; 1.
DR TIGRfams; TIGR00766; TIGR00766; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 337 AA; 37911 MW; C41B2A224902E311 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 337;
Best Local Similarity 71.4%; Pred. No. 31;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      10 IIFWIFW 16
Db      244 LFFWIFW 250

RESULT 14
Y147_MYCGE
ID Y147_MYCGE STANDARD; PRT; 375 AA.
AC P47393;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG147.
GN MG147.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,

```

```

RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39695; AAC71365.1; -.
DR TIGR; MG147; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 66 86 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 203 223 POTENTIAL.
FT TRANSMEM 234 254 POTENTIAL.
FT TRANSMEM 289 309 POTENTIAL.
FT TRANSMEM 338 358 POTENTIAL.
SQ SEQUENCE 375 AA; 43188 MW; A1AF07D574E8046 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 375;
Best Local Similarity 55.6%; Pred. No. 34;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      8 DFIFWIFW 16
Db      239 DFLVLWIFW 247

RESULT 15
DCEI_RAT
ID DCEI_RAT STANDARD; PRT; 593 AA.
AC P18088;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Glutamate decarboxylase, 67 kDa isoform (EC 4.1.1.15) (GAD-67)
DE (67 kDa glutamic acid decarboxylase).
GN GAD1 OR GAD67.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91014554; PubMed=2170798;
RA Wyborski R.J., Bond R.W., Gottlieb D.I.;
RT "Characterization of a cDNA coding for rat glutamic acid
decarboxylase."
RL Brain Res. Mol. Brain Res. 8:193-198(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90132703; PubMed=2299361;
RA Julien J.F., Samama P., Mallet J.;
RT "Rat brain glutamic acid decarboxylase sequence deduced from a cloned
cDNA."
RL J. Neurochem. 54:703-705(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92020930; PubMed=1924335;
RA Michelsen B.K., Petersen J.S., Boel E., Moldrup A., Dyrberg T.,
RA Madsen O.D.;
RT "Cloning, characterization, and autoimmunity recognition of rat islet
glutamic acid decarboxylase in insulin-dependent diabetes mellitus."
RT

```

```
Proc. Natl. Acad. Sci. U.S.A. 88:8754-8758(1991).
-!- FUNCTION: CATALYZES THE PRODUCTION OF GABA.
-!- CATALYTIC ACTIVITY: L-glutamate = 4-aminobutanoate + CO(2).
-!- COFACTOR: PYRIDOXAL PHOSPHATE.
-!- SUBUNIT: HOMODIMER.
-!- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND TYRDC).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or\_send\_an\_email\_to\_license@isb-sib.ch).
-----
DR EMBL; M34445; AAC42037.1; -
DR EMBL; X57572; CAA40800.1; -
DR EMBL; X57573; CAA40801.1; -
DR EMBL; M76177; AAA41184.1; -
DR PIR; A41367; A41367.
DR PIR; A43756; A43756.
DR InterPro; IPR002129; Pyridoxal_dec.
DR Pfam; PF00282; Pyridoxal_dec; 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
KW Neurotransmitter biosynthesis; Lyase; Decarboxylase;
KW Pyridoxal phosphate; Multigene family.
FT BINDING 404 404 PYRIDOXAL PHOSPHATE (POTENTIAL).
FT CONFLICT 103 103 L -> V (IN REF. 2).
FT CONFLICT 284 284 F -> S (IN REF. 2).
FT CONFLICT 287 288 EH -> AD (IN REF. 2).
FT CONFLICT 344 345 AG -> EA (IN REF. 2).
FT CONFLICT 347 347 T -> I (IN REF. 2).
FT CONFLICT 352 353 FD -> LE (IN REF. 2).
FT CONFLICT 380 380 L -> R (IN REF. 2).
SQ SEQUENCE 593 AA; 66640 MW; EF83239C30301F69 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 593;
Best Local Similarity 42.9%; Pred. No. 52;
Matches 6; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

Qy 5 CG--IDFIIFWIFW 16
Db 454 CGRHVDIFKFWLMW 467

Search completed: July 9, 2003, 15:08:58
Job time : 11.9823 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 7.22124 Seconds

(without alignments)
456.536 Million cell updates/sec

Title: US-09-854-133-587

Perfect score: 98

Sequence: 1 FOANGGIDFIIFWIFW 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_virus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB	ID	Description
1	49	50.0	495	1	Q9P9F4	Q9P9F4	methanosarc
2	49	50.0	495	17	Q8TQNS	Q8TQNS	methanosarc
3	47	48.0	275	8	Q9TDU3	Q9TDU3	echinococcu
4	47	48.0	491	2	Q9EV68	Q9EV68	rhizobium m
5	47	48.0	491	16	Q92YN6	Q92YN6	rhizobium m
6	47	48.0	535	8	Q958U2	Q958U2	echinococcu
7	46	46.9	192	5	Q8T4L1	Q8T4L1	brugia mala
8	46	46.9	310	11	Q8VEV1	Q8VEV1	mus musculu
9	46	46.9	530	8	Q953M3	Q953M3	echinococcu
10	46	46.9	3223	2	Q9RPH1	Q9RPH1	escherichia
11	46	46.9	3223	2	Q9RM48	Q9RM48	escherichia
12	46	46.9	3223	2	Q8VQR2	Q8VQR2	escherichia
13	46	46.9	3223	2	Q8VNR6	Q8VNR6	escherichia
14	45	45.9	269	5	Q968Z8	Q968Z8	plasmodium
15	45	45.9	319	11	Q91ZB6	Q91ZB6	mus musculu
16	45	45.9	343	4	Q96AM1	Q96AM1	homo sapien

17	45	45.9	343	11	Q8VCJ6	Q8VCJ6	mus musculu
18	45	45.9	856	5	Q9NA34	Q9NA34	caenorhabdi
19	44	44.9	149	8	Q02708	Q02708	podospora a
20	44	44.9	179	16	Q31816	Q31816	bacillus su
21	44	44.9	391	16	Q97L01	Q97L01	clostridium
22	44	44.9	531	5	Q9U8R4	Q9U8R4	drosophila
23	44	44.9	899	5	Q95XS9	Q95XS9	caenorhabdi
24	44	44.9	1440	5	Q9V3S5	Q9V3S5	drosophila
25	43	43.9	167	4	Q96DC5	Q96DC5	homo sapien
26	43	43.9	179	4	Q9UHQ2	Q9UHQ2	homo sapien
27	43	43.9	221	4	Q9UHQ3	Q9UHQ3	homo sapien
28	43	43.9	221	4	Q9BRV3	Q9BRV3	homo sapien
29	43	43.9	221	6	Q95KW8	Q95KW8	papio anubi
30	43	43.9	284	16	Q8ZBI8	Q8ZBI8	yersinia pe
31	43	43.9	332	16	Q9PMA3	Q9PMA3	campylobact
32	43	43.9	475	16	Q9ZML7	Q9ZML7	helicobacte
33	43	43.9	503	17	Q9HJK7	Q9HJK7	thermoplas
34	43	43.9	516	17	Q979N1	Q979N1	thermoplas
35	43	43.9	519	16	Q25661	Q25661	helicobacte
36	43	43.9	530	8	Q9B8V3	Q9B8V3	schistosoma
37	43	43.9	805	10	Q9FJ20	Q9FJ20	arabidopsis
38	42	42.9	194	16	Q8YPS8	Q8YPS8	anabaena sp
39	42	42.9	256	10	Q9SB30	Q9SB30	arabidopsis
40	42	42.9	351	5	Q76204	Q76204	schistosoma
41	42	42.9	386	3	Q93875	Q93875	candida alb
42	42	42.9	483	8	Q36097	Q36097	thellieria p
43	42	42.9	622	13	Q57661	Q57661	xenopus lae
44	42	42.9	671	10	Q9FMN5	Q9FMN5	arabidopsis
45	41	41.8	103	8	Q9G1S1	Q9G1S1	venereupis p

ALIGNMENTS

RESULT 1

Q9P9F4
ID Q9P9F4 PRELIMINARY; PRT; 495 AA.
AC Q9P9F4;
DC 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE F420H2 dehydrogenase subunit FpOM.
GN FpOM.
OS Methanosarcina mazel (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GOL;
RA Akken H.-J., Deppenmeier U.;
RT "purification and properties of an F420H2 dehydrogenase from
RT Methanosarcina mazel Gol.";
RL FEMS Microbiol. Lett. 154:231-237(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GOL;
RX MEDLINE=20309738; PubMed=10751389;
RA Baumer S., Ide T., Jacobi C., Johann A., Gottschalk G.,
RA Deppenmeier U.;
RT "The F420H2 Dehydrogenase from Methanosarcina mazel Is a Redox-driven
RT Proton Pump Closely Related to NADH Dehydrogenases.";
RL J. Biol. Chem. 275:17968-17973(2000).
DR EMBL; AF228525; AAF65740.1; .
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001750; Oxidored_q1.
DR Pfam; PF00361; oxidored_q1; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW NAD: Oxidoreductase.
SQ SEQUENCE 495 AA; 53972 MW; F288A37630CAFA00 CRC64;

Query Match 50.0%; Score 49; DB 1; Length 495;
Best Local Similarity 60.0%; Pred. No. 7.9;

Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFIIFWIFW 16
:|:|:|:|:|
Db 128 LDFVVFYIFW 137

RESULT 2

OS Q8TQNS PRELIMINARY; PRT; 495 AA.
AC Q8TQNS;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE F(420)H(2) dehydrogenase, subunit pfom.
GN FPOH OR MA1505.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity";
RL Genome Res. 12:532-542(2002);
DR EMBL; AE010819; AAM04919.1; -
KW Complete proteome.
SQ SEQUENCE 495 AA; 54138 MW; 278918B16BA7BAE6 CRC64;

Query Match 50.0%; Score 49; DB 17; Length 495;
Best Local Similarity 60.0%; Pred. No. 7.9;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFIIFWIFW 16
:|:|:|:|:|
Db 128 LDFVVFYIFW 137

RESULT 3

OS Q9TDU3 PRELIMINARY; PRT; 275 AA.
AC Q9TDU3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).
GN Col.
OS Echinococcus granulosus.
OG Echinodermata.
OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;
OC Cyclophyllidae; Taeniidae; Echinococcus.
OX NCBI_TaxID=6210;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=21036605; PubMed=11163447;
RA Nakao M., Sako Y., Yokoyama N., Fukunaga M., Ito A.;
RT "Mitochondrial genetic code in cestodes";
RL Mol. Biochem. Parasitol. 111:415-424(2000).
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-

r 3

CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 4 FERRICYTOCHROME
CC C + 2 H(2)O.
CC -!- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AB033407; BAA85324.1; -
DR InterPro; IPR000883; COX1.
DR Pfam; PF00115; COX1; 1.
DR PRINTS; PR01165; CYCOXIDASE1.
DR PROSITE; PS00077; COX1; UNKNOWN.1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 275 AA; 30298 MW; 20B1917902CB529D CRC64;

Query Match 48.0%; Score 47; DB 8; Length 275;
Best Local Similarity 50.0%; Pred. No. 9.5;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FOAGGIDFIIF 12
:|:|:|:|:|
Db 36 FSSSCGVDFLMF 47

RESULT 4

OS Q9EV68 PRELIMINARY; PRT; 491 AA.
AC Q9EV68;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Putative NADH-ubiquinone oxidoreductase subunit.
GN NUOM3.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RM41;
RA Putnoky P., Jady B., Chellapilla K.P., Barta F., Kiss E.;
RT "Sinorhizobium meliloti carries two sets of nuo genes";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AJ245399; CAC14150.1; -
DR InterPro; IPR003918; NADH_oxred4.
DR InterPro; IPR001750; Oxidored_q1.
DR Pfam; PF00361; oxidored_q1; 1.
DR PRINTS; PR01437; NUOXDRDTASE4.
KW NAD; Oxidoreductase; Transmembrane; Ubiquinone.
SQ SEQUENCE 491 AA; 52714 MW; A2C62574220C4EEA CRC64;

Query Match 48.0%; Score 47; DB 2; Length 491;
Best Local Similarity 41.7%; Pred. No. 16;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 5 CGIDFIIFWIFW 16
:|:|:|:|:|
Db 127 CALDLFLFYFVW 138

RESULT 5

Q92YN6 PRELIMINARY; PRT; 491 AA.
ID Q92YN6

DR	InterPro: IPR000893; COX1.
DR	Pfam: PF00115; COX1; 1.
DR	PRINTS; PR01165; CYCOXIDASE1.
DR	PROSITE; PS00077; COX1; UNKNOWN_1.
KW	Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW	Respiratory chain; Transmembrane.
SQ	SEQUENCE 535 AA; 60100 MW; 49944A4882F4FCED CRC64;
Query Match	48.0%; Score 47; DB 8; Length 535;
Best Local Similarity	50.0%; Pred. No. 18;
Matches	6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
OY	1 FOANCGDIFIIF 12 : : :
Dd	138 FSSCGVDLFMF 149
RESULT 7	
Q8T4L1	PRELIMINARY; PRT; 192 AA.
ID	Q8T4L1
AC	Q8T4L1;
DT	01-JUN-2002 (TrEMBLrel. 21, Created)
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	Thioredoxin.
GN	TRX.
OS	Brugia malayi.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC	Onchocercidae; Brugia.
OX	NCBI_TaxID=6279;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Wu Y., Bianco A.E.;
RT	"Thioredoxin of human filarial parasite Brugia malayi.";
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR	EMBL: AY080907; AAL91107.1; -
SQ	SEQUENCE 192 AA; 22071 MW; 617B22FA86CEA01 CRC64;
Query Match	46.9%; Score 46; DB 5; Length 192;
Best Local Similarity	50.0%; Pred. No. 9.8;
Matches	7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
OY	3 ANCGDIFIIFWIFW 16 : :
Dd	145 AYGCLNEIIYWVVF 158
RESULT 8	
Q8VEV1	PRELIMINARY; PRT; 310 AA.
ID	Q8VEV1
AC	Q8VEV1;
DT	01-MAR-2002 (TrEMBLrel. 20, Created)
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	Olfactory receptor MOR261-11.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Zhang X., Firestein S.J.;
RT	"The olfactory receptor gene superfamily of the mouse.";
RL	Nat. Neurosci. 0:0-0(2002).
RL	[2]
RN	SEQUENCE FROM N.A.
RP	Adams M.;
RA	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RL	EMBL: AY073797; AAL61460.1; -
DR	InterPro: IPR000276; GPCR_Rhodopsn.
DR	Pfam: PF00001; 7tm_1; 1.
DR	PRINTS; PR00237; GPCRRHODOPSN.
DR	PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; UNKNOWN_1.

DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Efal.
 GN Efal.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=83/39;
 RX MEDLINE=21192570; PubMed=11295191;
 RA Tauschek M., Strugnell R.A., Robins-Browne R.M.;
 RT "Characterization of the LEE pathogenicity islands of rabbit
 RT enteropathogenic Escherichia coli.";
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF453441; AAL57562.1;
 DR InterPro: IPR001917; NHtransf_2.
 DR InterPro: IPR000169; SHprot_acsite.
 DR PROSITE: PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
 DR PROSITE: PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
 SQ SEQUENCE 3223 AA; 365790 MW; E97D10B98FAD5658 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 3223;
 Best Local Similarity 43.8%; Pred. No. 1.4e+02;
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 FOANGCIDFIIFWIFW 16
 Db 2163 YQANTEIYLSNWIW 2178

RESULT 13
 Q8VNR6 PRELIMINARY; PRT; 3223 AA.
 AC Q8VNR6;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Efal-Lifa-Tox protein.
 GN Efal-Lifa-Tox
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=413/89-1;
 RA Benkel P., Chakraborty T.;
 RT "Genetic organisation and sequence of the LEE II locus in Shiga toxin-
 RT producing Escherichia coli.";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ277443; CAC81883.1;
 DR EMBL: AJ277443; CAC81883.1;
 DR InterPro: IPR001917; NHtransf_2.
 DR InterPro: IPR000169; SHprot_acsite.
 DR PROSITE: PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
 DR PROSITE: PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
 SQ SEQUENCE 3223 AA; 365566 MW; 28EBB4374FAB8FC7 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 3223;
 Best Local Similarity 43.8%; Pred. No. 1.4e+02;
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 FOANGCIDFIIFWIFW 16
 Db 2163 YQANTEIYLSNWIW 2178

RESULT 14
 Q96828 PRELIMINARY; PRT; 269 AA.
 AC Q96828;
 DT 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Ookinete surface protein Pos28-2.
 GN POS28-2.
 OS Plasmodium ovale (malaria parasite P. ovale).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36330;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NIGERIAN I/CDC;
 RX MEDLINE=21192570; PubMed=11295191;
 RA Tachibana M., Tsuboi T., Templeton T.J., Kaneko O., Torii M.;
 RT "presence of three distinct ookinete surface protein genes, Pos25,
 RT Pos28-1, and Pos28-2, in Plasmodium ovale.";
 RL Mol. Biochem. Parasitol. 113:341-344(2001).
 DR EMBL: AB051633; BAB43950.1;
 DR InterPro: IPR000561; EGF-like.
 DR SMART: SM00181; EGF; 3.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
 SQ SEQUENCE 269 AA; 29928 MW; 7A949BB42F036EEB CRC64;

Query Match 45.9%; Score 45; DB 5; Length 269;
 Best Local Similarity 46.2%; Pred. No. 19;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 FOANGCIDFIIFW 13
 Db 92 FRCNTRNVIFW 104

RESULT 15
 Q91ZB6 PRELIMINARY; PRT; 319 AA.
 ID Q91ZB6;
 AC Q91ZB6;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE G protein-coupled receptor (Fragment).
 GN MRGF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX MEDLINE=21435808; PubMed=11551509;
 RA Dong X., Han S.-K., Zylka M.J., Simon M.I., Anderson D.J.;
 RT "A Diverse Family of GPCRs Expressed in Specific Subsets of
 RT Nociceptive Somatosensory Neurons.";
 RL Cell 106:619-632(2001).
 DR EMBL: AY042211; AAK91802.1;
 DR InterPro: IPR000276; GPCR_Rhodpsn.
 DR Pfam: PF00001; 7tm_1; 1.
 DR PROSITE: PS00337; G_PROTEIN_RECEP_F1_1; UNKNOWN_1.
 DR PROSITE: PS0262; G_PROTEIN_RECEP_F1_2; 1.
 KW Receptor.
 FT NON_TER.
 SQ SEQUENCE 319 AA; 35955 MW; D1860538BA7A965C CRC64;

Query Match 45.9%; Score 45; DB 11; Length 319;
 Best Local Similarity 60.0%; Pred. No. 23;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15
 Db 232 GIDWFLFWVF 241

Search completed: July 9, 2003, 15:06:33
 Job time : 10.2212 secs

